

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OK protein - protein search, using sw model

Run on: April 20, 2004, 23:12:55, Search time 7.88961 Seconds

(without alignment)  
548.648 Million cell updates/sec

Title: US-09-816-989A-2

Perfect score: 213  
Sequence: 1 AKKYAKKAKKAKKAKKAA.....AKYKAAKAKKAAKAAKAA 45

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database:

PIR 78:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103.5	48.6	210	2 A25550	histone H1 - sea u
2	98	46.0	394	2 F90725	membrane spanning
3	98	46.0	394	2 G85576	membrane spanning
4	96	46.0	421	2 JY0057	tolA protein - Esc
5	94.5	44.4	239	2 AE1317	hypothetical prote
6	90	42.3	243	2 AE1689	hypothetical prote
7	90	42.3	1390	2 S51364	sperm tail-specific
8	89.5	42.0	211	2 A28100	histone H1-beta, e
9	89.5	42.0	214	2 G70673	probable hupB - My
10	88	41.3	899	2 B87553	DNA topoisomerase
11	86.5	40.6	384	2 B43582	outer membrane pro
12	86	40.4	199	2 F70742	hypothetical prote
13	86	40.4	344	2 S94153	mat101-1 protein -
14	86	39.4	347	2 E83525	tolA protein PA097
15	85	39.9	182	2 S61926	histone H1 homolog
16	82	38.5	248	1 H5URIP	histone H1, gonada
17	82	38.5	262	2 T35503	1-acylglycerol-3-P
18	82	38.5	566	2 T39675	asparaginyl-tRNA s
19	81.5	38.3	311	2 T17688	hypothetical prote
20	80.5	37.8	277	2 T34625	probable NUP/P60 f
21	80.5	37.8	445	2 T50972	probable nucleoti
22	80.5	37.8	952	2 T36664	probable DNA copo
23	80	37.6	194	1 H5TRI	histone H1 - trout
24	80	37.6	206	1 H5TRI	histone H1 - rainb
25	80	37.6	291	2 A81487	hypothetical prote
26	79.5	37.3	217	2 A26721	histone H1-gamma,
27	79	37.1	62	2 T30977	hypothetical prote
28	79	37.1	205	2 S19114	cgr-1 protein - C
29	79	37.1	244	2 S40436	histone H1 - midge

30	79	37.1	347	2 A81794	probable rotamase
31	78	36.6	217	2 S28309	hypothetical prote
32	78	36.6	231	2 S53589	histone H1 - Chlam
33	78	36.6	287	2 S45662	histone H1 - tomat
34	78	36.6	309	2 G83013	polyhydroxyalkanoa
35	77.5	36.4	376	2 AG0592	tolA protein (limp
36	77	36.2	220	2 A28456	histone H1.10 - ch
37	77	36.2	226	1 S51660	histone H1-5 [vali
38	77	36.2	523	2 B38145	invariant surface
39	77	36.2	580	2 B70868	probable transiera
40	76	35.7	229	2 I51227	histone H1A - Atri
41	76	35.7	241	2 JN0748	histone H1-II - Vo
42	76	35.7	388	2 AC0138	tolA colicin impor
43	75.5	35.4	220	2 JC5954	ribosomal protein
44	75.5	35.4	924	2 T06636	hypothetical prote
45	75	35.2	212	2 A28470	histone H1 - mouse

#### ALIGNMENTS

##### RESULT 1

A25550 histone H1 - sea urchin (lytechinus pictus)  
C/Species: Lytechinus pictus (painted urchin)  
C/Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 23-Jul-1999

C/Accession: A25550

R/Knowledge: J.A.; Childs, G.J.  
Nucleic Acids Res. 14, 8121-8133, 1986

A/Title: Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus e

A/Reference number: A25550; MUID:87040778; PMID:3022245

A/Accession: A25550

A/Molecule type: DNA

A/Residues: 1-210 <XNO>

A/Cross-references: GB:X04488; NID:G9616; PIDN:CAA28177.1; PID:G9617

C/Superfamily: histone H1

C/Keywords: chromosomal protein; DNA binding; nucleosome; nucleus

Query Match

Best Local Similarity 48.6%; Score 103.5; DB 2; Length 210;

Matches 28; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

DB

1 AKKYAKKAKKAKKAKKAAKAAKAAKAAKAAKAAKAA 42

166 AKKAKKPAKKPAKKAKKPAKKAKKPAKKAKKPAKKAKKAA 208

RESULT 2

F90725

membrane spanning protein TolA (imported) - Escherichia coli (strain O157:H7, substrain

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C/Accession: F90725

R/Knowledge: T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.C

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ger

A/Reference number: A59629; MUID:21156231; PMID:11258796

A/Accession: F90725

A/Molecule type: DNA

A/Residues: 1-394 <HAY>

A/Cross-references: GB:BA00007; PIDN:BA834197.1; PID:G13360233; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RIMD 0509952

C/Genetics:

A/Genes: EC80774

Query Match

Best Local Similarity 46.0%; Score 98; DB 2; Length 394;

Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

DB

2 KKYAKKAKKAKKAKKAAKAAKAAKAAKAAKAAKAA 45

166 AKKAKKPAKKPAKKAKKPAKKAKKPAKKAKKPAKKAKKAA 208







05-03-403-143A-1

```

; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co., Ltd.
; TITLE OF INVENTION: GLAISTRAMER ACETATE MOLECULAR WEIGHT MARKERS
; FILE REFERENCE: 60807-A

```

```

; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
;

```

	Query Match	42.7%	Score 91;	DB 4;	Length 223;
	Best Local Similarity	59.6%;	Pred. No.	0.0074;	
	Matches	28;	Conservative	5;	Mismatches 12; Indels 2; Gaps 2;
Oy	1 AKKYA-KKAAVEKKAKYKAAAEKKAKAYEKAA-AEKKAAKEAYEA	45			
	:               :				





```

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5990
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5990

```

```

Query Match      38.3%; Score 81.5; DB 4; Length 497;
Best Local Similarity 47.1%; Pred. No. 0.17;
Matches 24; Conservative 7; Mismatches 13; Indels 7; Gaps 1;

```

```

QY      2 KKVAKKAKAK-----AKKAYKAAEKKAKEKKAKEKAAKAAKEAAIEA 45
Db      236 KKEAKELAEQARQRAAKAEQAAQAQAAQKAAAEQAKATKAAKAA 286

```

```

RESULT 15
US-09-252-991A-32957
; Sequence 32957, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32957
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32957

```

```

Query Match      36.6%; Score 78; DB 4; Length 316;
Best Local Similarity 54.8%; Pred. No. 0.26;
Matches 23; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```

```

QY      1 AKKYAKKAKAKKAYKAAEAEEKKAAKYEKAAEKAAKEAA 42
Db      157 AKAAKPAKPAKPAKPAKTAAPAAKPAKPAKAAKPAKPAK 198

```

Search completed: April 20, 2004, 23:19:07  
 Job time : 9.35065 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

Run on: April 20, 2004, 23:08:59 ; Search time 6.13636 Seconds  
(without alignments)  
381.848 Million cell updates/sec

```

Title:      US-09-816-989A-2
Perfect Score: 213
Sequence:   1 AKKYAKKAAEKAKKAYKAA.....AKYEKAAAEKAAAEAYEA 45

```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	103.5	48.6	210	1	H1_LYPTI	P06114 lytehinus
2	98	46.0	421	1	TOLA_ECOLI	P19314 escherichia
3	90	42.3	1391	1	MS22_DROXY	O08656 dirosophila
4	89.5	42.0	208	1	DBH_MYCSM	O29b5 mycobacter
5	89.5	42.0	211	1	H1B_STRBU	P15869 strongylococ
6	89.5	42.0	214	1	DBH_MYCBO	O29b18 mycobacter
7	89.5	42.0	214	1	DBH_MYCTU	P95109 mycobacter
8	88.5	41.5	275	1	MS23_DROXY	O01355 dirosophila
9	86.5	40.6	384	1	TMPE_TREPH	P29720 treponema f
10	86.5	40.4	168	1	RS16_COREP	O8fp20 corynebacter
11	86	40.4	168	1	HHHA_MYCTU	O11142 mycobacter
12	86	40.4	344	1	MS21_DROXY	O08655 dirosophila
13	86	40.4	347	1	TOLA_PSEAE	P20600 pseudomonas
14	82	38.5	165	1	RS16_CORGL	O8mnx3 corynebacter
15	82	38.5	248	1	H1_PARAN	P02256 paracinthus
16	81	38.0	235	1	H1E_CHIPA	P02622 carchinonus
17	80.5	37.8	952	1	TOPI_STRCO	O9x969 streptomyces
18	80	37.6	194	1	H1_STRTR	P02252 salmo trutta
19	80	37.6	206	1	H1_ONCNY	P06350 oncorhynchus
20	79.5	37.3	217	1	H1G_STRUD	P07796 strongylococ
21	79	37.1	244	1	H1O_CHIRH	O07124 chironomus
22	78.5	36.9	236	1	H1_NEUCR	O8f0u2 neurospora
23	78.5	36.9	237	1	H1E_CHIRE	P40278 chironomus
24	78	36.6	233	1	H11_GLYVA	P40265 glyptotend
25	78	36.6	287	1	H11_GLYVA	P37228 glyptotend
26	77.5	36.4	299	1	H1_LYCES	P37228 lycopersicoc
27	77	36.2	219	1	RL22_DROME	P50857 dirosophila
28	77	36.2	225	1	H11O_CHICK	P08266 gallus galli
29	76.5	35.9	139	1	H15_HUMAN	P16401 homo sapien
30	76.5	35.9	232	1	ASR_KLEPN	O93mm5 klebsiella
31	76.5	35.9	233	1	H11_GLYVA	P40264 glyptotend
32	76.5	35.9	238	1	H11_GLYVA	P40263 glyptotend
33	76	35.7	240	1	H1_TREBAT	P27806 tritricum aet
					H12_VOLGA	O08865 volvox car

34	RS16_BACIN	184	1	O97915	bacteroides
35	75.5	212	1	RS14_HUMN	P50914
36	75	35.4	1	HL2_MOUSE	homo sapien
37	74.5	35.2	211	1	P15864
38	74.5	35.0	101	1	mus musculus
39	74.5	35.0	105	1	O93nm6
40	74.5	35.0	102	1	enterobacter
41	74	34.7	153	1	P58122
42	74	34.7	153	1	caulobacter
43	74	34.7	153	1	P39720
44	74	34.7	153	1	bacteriorhyme
45	74	34.7	153	1	HL2_MOUSE
46	74	34.7	153	1	HL2_MOUSE
47	74	34.7	153	1	HL2_MOUSE
48	74	34.7	153	1	HL2_MOUSE
49	74	34.7	153	1	HL2_MOUSE
50	74	34.7	153	1	HL2_MOUSE
51	74	34.7	153	1	HL2_MOUSE
52	74	34.7	153	1	HL2_MOUSE
53	74	34.7	153	1	HL2_MOUSE
54	74	34.7	153	1	HL2_MOUSE
55	74	34.7	153	1	HL2_MOUSE
56	74	34.7	153	1	HL2_MOUSE
57	74	34.7	153	1	HL2_MOUSE
58	74	34.7	153	1	HL2_MOUSE
59	74	34.7	153	1	HL2_MOUSE
60	74	34.7	153	1	HL2_MOUSE
61	74	34.7	153	1	HL2_MOUSE
62	74	34.7	153	1	HL2_MOUSE
63	74	34.7	153	1	HL2_MOUSE
64	74	34.7	153	1	HL2_MOUSE
65	74	34.7	153	1	HL2_MOUSE
66	74	34.7	153	1	HL2_MOUSE
67	74	34.7	153	1	HL2_MOUSE
68	74	34.7	153	1	HL2_MOUSE
69	74	34.7	153	1	HL2_MOUSE
70	74	34.7	153	1	HL2_MOUSE
71	74	34.7	153	1	HL2_MOUSE
72	74	34.7	153	1	HL2_MOUSE
73	74	34.7	153	1	HL2_MOUSE
74	74	34.7	153	1	HL2_MOUSE
75	74	34.7	153	1	HL2_MOUSE
76	74	34.7	153	1	HL2_MOUSE
77	74	34.7	153	1	HL2_MOUSE
78	74	34.7	153	1	HL2_MOUSE
79	74	34.7	153	1	HL2_MOUSE
80	74	34.7	153	1	HL2_MOUSE
81	74	34.7	153	1	HL2_MOUSE
82	74	34.7	153	1	HL2_MOUSE
83	74	34.7	153	1	HL2_MOUSE
84	74	34.7	153	1	HL2_MOUSE
85	74	34.7	153	1	HL2_MOUSE
86	74	34.7	153	1	HL2_MOUSE
87	74	34.7	153	1	HL2_MOUSE
88	74	34.7	153	1	HL2_MOUSE
89	74	34.7	153	1	HL2_MOUSE
90	74	34.7	153	1	HL2_MOUSE
91	74	34.7	153	1	HL2_MOUSE
92	74	34.7	153	1	HL2_MOUSE
93	74	34.7	153	1	HL2_MOUSE
94	74	34.7	153	1	HL2_MOUSE
95	74	34.7	153	1	HL2_MOUSE
96	74	34.7	153	1	HL2_MOUSE
97	74	34.7	153	1	HL2_MOUSE
98	74	34.7	153	1	HL2_MOUSE
99	74	34.7	153	1	HL2_MOUSE
100	74	34.7	153	1	HL2_MOUSE
101	74	34.7	153	1	HL2_MOUSE
102	74	34.7	153	1	HL2_MOUSE
103	74	34.7	153	1	HL2_MOUSE

## ALIGNMENTS

```

RESULT 1
HL HYPI
ID_ H1 LYPI STANDARD; PR; 210 AA.
AC P06144;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Late histone H1.
OS Lyechnus pictus (Painted sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinodae; Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;
OC Lyechnus.
OC NCBI_TaxID=7653;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=87040778; PubMed=3022245;
RA Knowles J.A., Childs G.J.;
RT "Comparison of the late H1 histone genes of the sea urchins
RL Lyechnus pictus and Strongylocentrotus purpuratus."
RL Nucleic Acids Res. 14:8121-8133(1986).
CC -1- FUNCTION: Histones H1 are necessary for the condensation of
CC nucleosome chains into higher order structures.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the histone H1/H5 family.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X04488; CAA28177.1; -.
CC DR PIR; A25550; A25550.
CC DR HSSP; P02259; 1HST.
CC DR InterPro; IPR005818; Histone_H1/H5.
CC DR InterPro; IPR005819; Histone_H5.
CC DR InterPro; IPR003216; Linkerhist_N.
CC DR Pfam; PF00538; linker_histone; 1.
CC DR PRINTS; PR00624; HISTONEN5.
CC DR ProDom; PD000373; linkerhist_N; 1.
CC DR SMART; SM00526; H15. 1.
CC KW Chromosomal protein, Nuclear protein; DNA-binding; Multigene family.
CC SEQUENCE 210 AA; 21746 MW; 08C3F6489400782 CRC64;
KW

```

RESULT 2	TOL_A_ECOLI	STANDARD;	PRT;	421 AA.
ID	TOL_A_ECOLI			
AC	P19934;			
DT	01-FEB-1991 (rel. 17, Created)			
DT	01-FEB-1991 (rel. 17, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	TOL_A protein.			
GN	TOL_A OR CIM OR EXCC OR LKY OR B0739.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN=JM105;			
RC	MEDLINE=90078104; PubMed=2687247;			
RX	Levengood S.K., Webster R.B.;			
RT	"Nucleotide sequences of the tolA and tolB genes and localization of			
RT	their products, components of a multistep translocation system in			
RT	Escherichia coli.";			
RL	J. Bacteriol. 171:6600-6609 (1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KL12 / MG1655;			
RC	MEDLINE=97426617; PubMed=9278503;			
RA	Blatner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.";			
RL	Science 277:1453-1474 (1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RP	SPRAIN=KL12;			
RC	MEDLINE=97061202; PubMed=8905232;			
RA	Oshina T., Alpa H., Babs T., Fujita K., Hayashi K., Honjo A.,			
RA	Iremoto K., Noda T., Itoh T., Kajihara M., Kanai K., Kasimoto K.,			
RA	Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,			
RA	Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.			
RA	Sampai G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,			
RA	Yano M., Horiuchi T.;			
RT	"A 718-bp DNA sequence of the Escherichia coli K-12 genome			
RT	corresponding to the 12.7-28.0 min region on the linkage map.";			
RL	DNA Res. 3:137-155 (1996).			
RN	[4]			
RP	DOMAINS.			
RP	MEDLINE=91296736; PubMed=2068069;			
RA	Levengood S.K., Beyer W.F. Jr., Webster R.E.;			
RT	"tolA: a membrane protein involved in colicin uptake contains an			
RT	extended helical region.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943 (1991).			
RN	[5]			
RP	INTERACTION WITH PORINS.			
RP	MEDLINE=97133271; PubMed=8978668;			
RA	Deronche R., Gavioli M., Beredetti H., Pilirov A., Lazdunski C.,			
RA	Floures R.;			
RT	"tolA central domain interacts with Escherichia coli porins.";			
RL	EMBO J. 15:6408-6415 (1996).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 296-421.			
RA	Lubkowksi J., Henneke F., Plueckthun A., Wlodawer A.;			
RT	"Immune phage infection: crystal structure of g3p in complex			
RT	with its coreceptor, the C-terminal domain of tolA.";			
RL	Structure 7:711-722 (1999).			
CC	-I- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A			
CC	COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE			
CC	COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL			
CC	BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLLOCATION			
CC	OF BACTERIOPHAGE DNA.			
CC	-I- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE			
CC	AND LAMB.			

```

CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outletation -
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/send_email_to_license@ebi.ac.uk).
CC -----
CC DR EMBL; M28332; AAA24683.1; .
CC DR EMBL; AE000177; AAC73853.1; .
CC DR EMBL; D90713; BAA35405.1; .
CC DR PIR; JY0057; JY0057.
CC DR PDB; 1TOL; 20-MAY-99.
CC DR EcGene; EG11007; tolA.
CC KW Transport; Protein transport; Bacteriocin transport; Transmembrane;
CC Repeat; Inner membrane; 3D-structure; Complete proteome.
CC FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
CC FT TRANSSEM 14 34 POTENTIAL.
CC FT DOMAIN 35 421 PERIPLASMIC (POTENTIAL).
CC FT DOMAIN 48 310 DOMAIN II (ALPHA-HELICAL).
CC FT DOMAIN 311 421 DOMAIN III (FUNCTIONAL).
CC FT DOMAIN 224 278 10 X TANDEN REPEATS OF [ED-K(1,2)-A(2,4)].
CC
CC FT DISULFID 363 388
CC FT HELIX 335 349
CC FT TURN 350 351
CC FT TURN 353 354
CC FT HELIX 355 358
CC FT TURN 359 360
CC FT STRAND 363 369
CC FT TURN 371 372
CC FT STRAND 375 383
CC FT HELIX 385 397
CC FT HELIX 406 412
CC FT TURN 413 414
CC FT STRAND 416 421
CC SQ SEQUENCE 421 AA; 43156 MW; 8B2F52BA4B97C655E CRC64;
CC
CC Query Match 46.0%; Score 98; DB 1; Length 421;
CC Best Local Similarity 65.3%; Pred. No. 0.0068;
CC Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;
CC
Ox 2 KTVAKKAKAEKA---KK-AVKAALKAKAEKAEKAEKAEKAEVAEA 45
Db 247 KKAEEKKAAEKXADKKAAEKAAADKKAAA-AAVAAEKAAAAKAAABA 294

```

```

CC -1- FUNCTION: Possible structural role in the sperm tail.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Testis. Primary spermatocytes and early
CC spermatids.
CC -1- DOMAIN: The predominant structure is alpha-helical.
CC -1- POLYMORPHISM: Length polymorphisms exist between different
CC strains, most likely caused by length variations within the tandem
CC repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X73481; CAAS1876.1; -.
DR PIR, S51364; S51364.
DR FlyBase; FBgn0020733; Dhyd\msr101(2).
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007288; P:axosome assembly; IEA.
KM Sperm; Repeat; Multigene family; Polymorphism.
FT DOMAIN 332 1268
SQ SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F4878 CRC64;

Query Match
Best Local Similarity 42.3%; Score 90; DB 1; Length 1391;
Matches 26; Conservative 4; Mismatches 11; Indels 4; Gaps 2;

QY 2 KKYAKKAKAEK-AKAYKAAEAKK---AAKYKAAEAKKAAEAA 42
663 KETKEXKXCKEAKKAKKKAKEKKKCAAKKKEKKAKEKKKCEEA 707

RESULT 4
DB MYCSM STANDARD; PRT; 208 AA.
AC Q9ZHC5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein HU homolog (Histone-like protein) [Hlp].
OS HUP OR HLP.
OC Mycobacterium smegmatis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 700084 / mc(2)155;
RX MEDLINE=99110209; PubMed=8949418;
RA Lee B.H., Munagasu-Oel B., Dick T.;
RT "Upregulation of a histone-like protein in dormant Mycobacterium
RT smegmatis.";
RL Mot. Genet. 260:475-479(1998).
CC -1- FUNCTION: Histone-like DNA-binding protein which is capable of
CC wrapping DNA to stabilize it, and thus to prevent its denaturation
CC under extreme environmental conditions (By similarity).
CC -1- SIMILARITY: Belongs to the bacterial histone-like protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF068138; AADI3809.1; -.
DR HSSP, P02346; LHU.

```

```

DR InterPro; IPR000119; Bac DNAbind.
DR Pfam; PF00216; Bac DNA binding; 1.
DR PRINTS; PR01727; DNABINDINGHU.
DR ProDom; PD000945; Bac DNABind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE LIKE; 1.
KM DNA-binding; DNA condensation; Repeat.
FT DOMAIN 1 90
FT DOMAIN 101 205
SQ SEQUENCE 208 AA; 21230 MW; C4F577F61F7EF09 CRC64;

Query Match
Best Local Similarity 42.0%; Score 89.5; DB 1; Length 208;
Matches 29; Conservative 3; Mismatches 13; Indels 9; Gaps 2;

QY 1 AKKYAKKAKAEK-----KKAYKAAAKKAAK--EKAAEQAQAAEAYEA 45
111 AKKAAKKA PAKKAAKKTATYAAAKKAPAKKATAPAKKAAATYAPAKKAAATKA 164

RESULT 5
DB MYCSM STANDARD; PRT; 211 AA.
AC P15869;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1-beta, late embryonic.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Echinoidea; Echinodidae; Echinacea; Echinodidae; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=88246461; PubMed=2837660;
RA Lai Z.-C., Childs G.;
RT "Characterization of the structure and transcriptional patterns of
RT the gene encoding the late histone subtyp H1-beta of the sea urchin
RT Strongylocentrotus purpuratus.";
RL Mol. Cell. Biol. 8:1842-1844(1988).
CC -1- FUNCTION: Histones H1 are necessary for the condensation of
CC nucleosome chains into higher order structures.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the histone H1/H5 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M20314; AAA30052.1; -.
DR PIR, A28100; A28100.
DR HSSP, P02259; IHST.
DR InterPro; IPR005818; Histone H1/H5.
DR InterPro; IPR005819; Histone H5.
DR InterPro; IPR003216; LinkerHist N.
DR Pfam; PF00538; Linker histone; I.
DR PRINTS; PR00624; HISTONH5.
DR ProDom; PD000373; LinkerHist_N; 1.
DR SMART; SM00526; H15; 1.
KM Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 211 AA; 22169 MW; 9F214581334BE7A CRC64;

Query Match
Best Local Similarity 42.0%; Score 89.5; DB 1; Length 211;
Matches 26; Conservative 3; Mismatches 10; Indels 3; Gaps 2;

QY 1 AKKYAKKAKAEKAKAYKAAEAKKAAKAAKAAEAKKAAEAA 42

```

Db 139 SKKTKKKKKPAKAKKPA-AKKA--KPAKPAKKA 177

RESULT 6  
DBH\_MYCBO STANDARD; PRT; 214 AA.

AC 09XB18; 0985J5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp)  
GN HUP OR HUP OR HUP OR MDP1 OR MB3010C.  
OS Mycobacterium bovis.  
OC Bacteria; Actinobacteria; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1765;

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[2]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[3]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[4]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[5]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[6]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[7]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[8]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[9]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[10]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[11]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[12]

Query Match 42.0%; Score 89.5; DB 1; Length 214;  
Best Local Similarity 55.6%; Pred. No. 0.025; Mismatches 13; Indels 3; Gaps 1;  
Matches 25; Conservative 4;

DBH\_MYCBO STANDARD; PRT; 214 AA.

AC 09XB18; 0985J5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp)  
GN HUP OR HUP OR HUP OR MDP1 OR MB3010C.  
OS Mycobacterium bovis.  
OC Bacteria; Actinobacteria; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1765;

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[2]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[3]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[4]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[5]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[6]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[7]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[8]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[9]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[10]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[11]

RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
RC STRAIN=AMS;  
RL "HUP-A target for differentiation of M. tuberculosis and M. bovis."  
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[12]

Best Local Similarity 54.5%; Pred. No. 0.083;



CC -1- DOMAIN: HEPARIN BINDING SEEMS TO REQUIRE THE C-TERMINAL  
 CC DOMAIN OF HBHA. PROGRESSIVE TRUNCATIONS FROM THE C-TERMINAL END  
 CC DIMINISH THE AFFINITY FOR HEPARIN.  
 CC -1- PM: GLYCOSYLATED. GLYCOSYLATION MAY PROTECT THE PROTEIN FROM  
 CC PROTEOLYTIC DEGRADATION AND BE IMPORTANT FOR HEMAGGLUTINATION. IT  
 CC SUGGESTS THAT THE CARBOHYDRATE MOIETY MAY BE LOCATED WITHIN THE C-  
 CC TERMINAL DOMAIN OF HBHA.  
 CC -1- MISCELLANEOUS: SERUM FROM PATIENTS DIAGNOSED WITH ACTIVE  
 CC TUBERCULOSIS THAT HAD NOT BEEN VACCINATED CONTAINS ANTIBODIES THAT  
 CC RECOGNIZE HBHA, WHEREAS SERUM FROM HEALTHY INDIVIDUALS DOES NOT  
 CC CONTAIN ANY.  
 CC -1- SIMILARITY: STRONG. TO M. LEPRAE HBHA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF074390; AAC26052.1; -  
 DR EMBL; Z77162; CAB00936.1; -  
 DR EMBL; AE006951; AAK44716.1; ALT\_INIT.  
 DR EMBL; BX248335; CAD93348.1; -  
 DR PIR; F70742; F70742.  
 DR TIGR; MT0493; -  
 DR TubercuList; Rv0475; -  
 KM Cell adhesion; Heparin-binding; Hemagglutinin; Glycoprotein;  
 KM Virulence; Complete proteome.  
 FT INIT\_MET 0  
 FT DOMAIN 151 193 ALA/LYS-RICH.  
 FT CONFLICT 120 120 R->P (IN REF. 1).  
 SQ SEQUENCE 198 AA; 21403 MM; 5137605F1EB6042 CRC64;  
 Query Match 40.4%; Score 86; DB 1; Length 198;  
 Best Local Similarity 65.6%; Pred. No. 0.052;  
 Matches 21; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 Oy 11 EKAKKAYKAAEKAKYKAAEKAKAKAA 42  
 Db 160 KKAAPAKKAPAKKAPAKKAPAKKAPAKKAA 191  
 RESULT 12  
 MS11 DROXY STANDARD; PRT; 344 AA.  
 AC 008695;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Axoneme-associated protein met101(1).  
 GN MS1101(1).  
 OS Drosophila hydei (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 ON NCBI\_TaxID=7224;  
 RX 11  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=Testis;  
 RA MEDLINE=94200512; PubMed=8150205;  
 RA Neesen U., Benemann H., Heinlein U.A.O.,  
 RA "The Drosophila hydei gene Dms1101(1) encodes a testis-specific,  
 RA repetitive, axoneme-associated protein with differential abundance in  
 RA X chromosome deletion mutant flies.";  
 RL Dev. Biol. 162:414-425(1994).  
 CC -1- FUNCTION: Possible structural role in the sperm tail. It is  
 CC associated with axonemal structures  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- TISSUE SPECIFICITY: Testis. Located in spermatocytes and  
 CC spermatid bundles.  
 CC -1- DOMAIN: The predominant structure is alpha-helical.

CC -1- POLYMORPHISM: The number of repeats varies between strains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X73480; CAA51875.1; -  
 DR PIR; S34153; S34153.  
 DR FlyBase; FBgn0011816; Dhyd\met101(1).  
 DR GO; GO:0005737; C:cytoplasm; IDA.  
 DR GO; GO:0001198; F:structural molecule activity; IEP.  
 DR GO; GO:0007288; P:axoneme assembly; IEP.  
 KW Sperm; Repeat; Multigene family.  
 FT DOMAIN 58 337  
 FT 19 X 16 AA APPROXIMATE TANDEM REPEATS OF  
 FT K-K-K-C-X-E-X-A-[KQ]-K-X-X-E-X-A-X.  
 SQ SEQUENCE 344 AA; 37793 MM; 24C65D2510387E2A CRC64;  
 Query Match 40.4%; Score 86; DB 1; Length 344;  
 Best Local Similarity 57.8%; Pred. No. 0.084;  
 Matches 26; Conservative 3; Mismatches 12; Indels 4; Gaps 2;  
 Oy 2 KKYAKKAK-AAEKAKYKAAKAKK---AAKYEKAAAEKAAKAA 42  
 Db 69 KEAAKKKCAKAAKKEKAAKKEKAAKKEKAAKKEKAAKCAEAA 113  
 RESULT 13  
 ID TOLA\_PSEAB STANDARD; PRT; 347 AA.  
 AC P50600;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Tola protein.  
 GN TOLA OR PA0971.  
 OS Pseudomonas aeruginosa.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 CC Pseudomonadaceae; Pseudomonas.  
 ON NCBI\_TaxID=287;  
 RX 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PAO;  
 RA MEDLINE=97113525; PubMed=8955385;  
 RA Denis J.U., Lafontaine E.R., Sokol P.A.,  
 RA "Identification and characterization of the tolQRA genes of  
 RA Pseudomonas aeruginosa.";  
 RL J. Bacteriol. 178:7059-7068(1996).  
 RN 12  
 RP REVISIONS TO N-TERMINUS.  
 RA Duan K., Sokol P.A.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN 13  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAO1;  
 RA MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Britkman P.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltry L., Tolentino E., Westbrock-Adams S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kae A., Ladwig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reller J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RA "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
 RA opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -1- FUNCTION: Involved in the tonB-independent uptake of proteins (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane  
 CC (Potential).







Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:51 11 November 2014

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OC NCBI\_Taxid=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=06:HI / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Weich R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:117020-117024(2002).  
 DR EMBL; AE016757; AAN79291.1; -  
 KM Complete proteome.  
 SQ SEQUENCE 421 AA; 43184 MM; DB296626F056D385 CRC64;  
 Query Match 46.0%; Score 98; DB 16; Length 421;  
 Best Local Similarity 65.3%; Pred. No. 0.051;  
 Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;  
 QY 2 KKYAKKAAEKAA--KK--AYKAAEKAAKYEKAAEKAAEKAAEAYEA 45  
 DB 247 KKAEEKAAEKAAADKKAEEKAAADKKA--AKAAEKAAAKAAEAAE 294  
 RESULT 6  
 ID Q83SA1 PRELIMINARY; PRT; 413 AA.  
 AC Q83SA1;  
 DT 01-JUN-2003 (TReMBLrel. 24, Created)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Membrane spanning protein, required for outer membrane integrity.  
 GN TOXA OR SPO558 OR S0571.  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Shigella.  
 OC NCBI\_Taxid=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=301 / Serotype 2a;  
 RX MEDLINE=22272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Zhang F., Zhang X., Yang J., Yang H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.,  
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
 RT through comparison with genomes of Escherichia coli K12 and O157,"  
 RL Nucleic Acids Res. 30:4432-4441(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.,  
 RT "Complete genome sequence and comparative genomics of Shigella  
 RT flexneri serotype 2a strain 2457T."  
 RL Infect. Immun. 71:2775-2786(2003).  
 DR EMBL; AE015086; AAN42202.1; -  
 DR EMBL; AE016797; AAP16075.1; -  
 KM Complete proteome.  
 SQ SEQUENCE 413 AA; 42355 MM; 93E10F2C5DE60DE8 CRC64;  
 Query Match 44.6%; Score 95; DB 16; Length 413;  
 Best Local Similarity 63.3%; Pred. No. 0.059;  
 Matches 31; Conservative 3; Mismatches 9; Indels 6; Gaps 3;  
 QY 2 KKYAKKAAEKAA--KK--AYKAAEKAAKYEKAAEKAAEKAAEAYEA 45

DB 239 KKAEEKAAADKKAADKKAEEKAAADKKA--AKAAEKAAAKAAEAAE 286  
 RESULT 7  
 ID Q8Y5W4 PRELIMINARY; PRT; 239 AA.  
 AC Q8Y5W4;  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DE Hypothetical protein lmo1941.  
 GN LMO1941.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OC NCBI\_Taxid=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EGD-e / Serovar 1/2a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaeser P., Frangoul L., Buchrieser C., Rusnock C., Amend A.,  
 RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Darvar A., Deboux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Desurget O.,  
 RA Eutlian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut U., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kretz U., Kuhn M., Kunst F., Kuzepkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordisiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Pucell R.,  
 RA Remmel B., Rose M., Schuener T., Simoes N., Tierrez P.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland U., Cossart P.,  
 RT "Comparative genomics of Listeria species,"  
 RL Science 294:849-852(2001).  
 DR EMBL; AL591981; CAD00019.1; -  
 DR PIR; AB1317; AE1317.  
 DR L1stlist; LMO1941; -  
 DR GO; GO:0016988; Pcell wall catabolism; IEA.  
 DR InterPro; IPR002482; LysM.  
 DR Pfam; PF01476; LysM; 1.  
 DR SMART; SM00257; LysM; 1.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 239 AA; 25836 MM; 72E59D57E0D7832 CRC64;  
 Query Match 44.4%; Score 94.5; DB 16; Length 239;  
 Best Local Similarity 56.2%; Pred. No. 0.066;  
 Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;  
 QY 2 KKYAKKAAEKAA--KK--AYKAAEKAAKYEKAAEKAAEKAAEAYEA 44  
 DB 124 KKAEEKAAADKKAADKKAEEKAAADKKA--AKAAEKAAAKAAEAAE 171  
 RESULT 8  
 ID Q81H69 PRELIMINARY; PRT; 248 AA.  
 AC Q81H69;  
 DT 01-JUN-2003 (TReMBLrel. 24, Created)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DE TcdA-E operon negative regulator.  
 GN BC0954.  
 OS Bacillus cereus (strain ATCC 14579 / DSM 31).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OC NCBI\_Taxid=226900;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=22608415; PubMed=12721630;  
 RX MEDLINE=22608415; PubMed=12721630;  
 RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candolon B.,  
 RA Kapatal V., Bhattacharya A., Resnik G., Mikhailova N., Lapidis A.,  
 RA Chu L., Mazur M., Goldsman E., Larsen N., D'Souza M., Watrous T.,  
 RA Grechkin Y., Fusch G., Haseikorn R., Forstner M., Ehrlich S.D.,  
 RA Overbeek R., Kyriides N.,  
 RT "Genome sequence of Bacillus cereus and comparative analysis with

```

RA  Huang C.H., Kieseer T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA  Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S., Taylor K.,
RA  Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA  Warren T., Weitzorek A., Woodward J., Barrett B.G., Parkhill J.,
RA  Hopwood D.A.,
RT  "Complete genome sequence of the model actinomycete Streptomyces
RL  coelicolor A3(2).";
RL  Nature 417:141-147(2002).
DR  EMBL, AL931110, CACC8545.1; -.
KW  Complete proteome.
SQ  SEQUENCE 244 AA; 25524 MW; 6199D62CA23A7B0 CRC64;

OY  5 AKKAAEKAKK--AKYAAEKKAAYKKAEEKAAEAAYE 44
Db  78 SQYVAIEAKKEAAKATKATKATYAEKKAAEKAAAKRAKE 119

RESULT 11
O7W3X2 PRELIMINARY; PRT; 197 AA.
AC O7W3X2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Histone protein.
GN BPH1 OR BP3905.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_Taxid=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; Pubmed=12910271;
RA Parkhill J., Sebatina M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Harraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Bauman D., Bacon N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamblin N., Hauser H., Holtjoy D., Jagels K.,
RA Leather S., Moulé S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitz E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skellon J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrett B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL, BX640435, CAC39188.1; -.
KW Complete proteome.
SQ SEQUENCE 197 AA; 19583 MW; 11D642F2BFA3E1EE CRC64;

OY 1 AKKYAKKAKK--AKKAYAAEKKAAYKKAEEKAAEAAYE 41
Db 60 AKKYAKKAYAKKPAKKYAKKAYAKKAAVA-KKAYAKKAYAKKA 101

Query Match 42.5%; Score 90.5; DB 16; Length 197;
Best Local Similarity 60.5%; Pred. No. 0.13;
Matches 26; Conservative 4; Mismatches 10; Indels 3; Gaps 2;

RESULT 12
O92A67 PRELIMINARY; PRT; 243 AA.
AC O92A67;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein lin2055.
RN LIN2055.

```

OS *Listeria innocua*.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIP 11262 / Serovar 6a;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Ruenliok C., Amend A.,  
BAguero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,  
RA Charbit A., Cherouari F., Couve E., de Daruvar A., Dehoux P.,  
RA Domian E., Dominguez-Bernal G., Duchaud E., Durant L., Duseux O.,  
RA Entian K.-D., Feihl H., Garcia-del Portillo F., Garrido P.,  
RA Gantier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkay G.,  
RA Madueno E., Matounam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tietze A.,  
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;  
RT "Comparative genomics of *Listeria species*.";  
RL Science 294:849-852(2001).  
DR EMBL; AL596170; CAC97285.1; -  
DR PIR; AE1689; AE1689.  
DR ListList; LTN02055; -  
DR GO; GO:0016998; P:cell wall catabolism; IEA.  
DR InterPro; IPR02482; LysM.  
DR Pfam; PF01476; LysM; 1.  
DR SMART; SM00257; LysM; 1.  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 243 AA; 25963 MW; 6B2493D43B159D1 CRC64;  
  
Query Match 42.3%; Score 90; DB 16; Length 243;  
Best Local Similarity 45.5%; Pred. No. 0.18;  
Matches 30; Conservative 3; Mismatches 11; Indels 22; Gaps 2;  
  
QY 1 AKKVA-KKAKAEKKAKKAAKAAKAA-----AKKAAKAAKAA 38  
DB 106 AKKAAEKAAEKAAEKAAEKAAEKAAADKKSQEDPAKAAKAAKAAKAAEKAAEKAAKAA 165  
QY 39 KEAAVE 44  
DB 166 DKAAKE 171  
  
RESULT 13  
Q8RL36 ID Q8RL36 PRELIMINARY; PRT; 182 AA.  
AC Q8RL36;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Histone-like protein Bph2 (Fragment).  
GN BPH2.  
OS Burkholderia multivorans.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Burkholderia.  
OX NCBI\_TaxID=87883;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kholit A., Plesa M., Cornelis P.;  
RT "Mini-Tnori, a new mini-transposon for in vivo protein epitope  
RT tagging.";  
RL Submitted (FEF-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY089703; AM009090.2; -  
FT NON\_TER 182  
SQ SEQUENCE 182 AA; 18768 MW; 243C5AECDA5A0234 CRC64;  
  
Query Match 41.8%; Score 89; DB 2; Length 182;  
Best Local Similarity 60.4%; Pred. No. 0.18;  
Matches 29; Conservative 3; Mismatches 10; Indels 6; Gaps 3;  
  
QY 1 AKKVA-KKAKAEKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 42  
DB 19 AKKVAOPAKKAAAVKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 66

RESULT 14  
Q9WMX1 ID Q9WMX1 PRELIMINARY; PRT; 372 AA.  
AC Q9WMX1;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Tola protein.  
GN TOLA.  
OS Pseudomonas putida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=mt-2;  
RX MEDLINE=96198174; PubMed=8626299;  
RA Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.;  
RT "The Pseudomonas putida peptidoglycan-associated outer membrane  
RT lipoprotein (PAL) is involved in maintenance of the integrity of the  
RT cell envelope.";  
RL J. Bacteriol. 178:1699-1706(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=mt-2;  
RA Ramos-Gonzalez I.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=mt-2;  
RA Rodriguez-Herva J.J.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=mt-2;  
RX MEDLINE=96422022; PubMed=8824639;  
RA Rodriguez-Herva J.J., Ramos J.;  
RT "Characterization of an Opl null mutant of *Pseudomonas putida*.";  
RL J. Bacteriol. 178:5836-5840(1996).  
DR EMBL; X74218; CAB50780.1; -  
DR GO; GO:000786; C:nucleosome; IEA.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0006334; P:nucleosome assembly; IEA.  
DR InterPro; IPR005819; Histone H5.  
DR PRINTS; PRO0624; HISTONEH5.  
DR TIGRFAms; TIGR01352; cons\_Cterm; 1.  
SQ SEQUENCE 372 AA; 40133 MW; 87F49785EC3C0BC CRC64;  
  
Query Match 41.5%; Score 88.5; DB 2; Length 372;  
Best Local Similarity 39.7%; Pred. No. 0.39;  
Matches 27; Conservative 8; Mismatches 10; Indels 23; Gaps 1;  
  
QY 1 AKKVA-KKAKAEKKAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 37  
DB 121 AEDAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 180  
QY 38 AEAAYEA 45  
DB 181 AEAAKKA 188  
  
RESULT 15  
Q88N16 ID Q88N16 PRELIMINARY; PRT; 372 AA.  
AC Q88N16;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Biopolymer transport protein Tola.

OS TOLA OR P1221.  
 SN Pseudomonas putida (strain KT2440).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID:160488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:22423060; PubMed:12534463;  
 RA Nelson K.E., Metnel C., Paulsen I.T., Dodson R.J., Hiltbert H.,  
 RA Mattins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
 RA Brinkbe L., Beaman M., DeBoy R.T., Daugherty S., Kolonay J.,  
 RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,  
 RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzaz A.,  
 RA Uetreacker T., Rizzo M., Lee K., Kosack D., Moesti D., Medler H.,  
 RA Lander U., Stjepandic D., Hohnsels U., Streletz M., Helm S.,  
 RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuenmler B.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative analysis of the  
 RT metabolically versatile Pseudomonas putida KT2440.";  
 RL Environ. Microbiol. 4:799-808(2002).  
 DR EMBL; AE016778; AAN66845.1; -.  
 DR TIGR; P1221; -.  
 DR GO; GO:0000786; C:nucleosome; IEA.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0006334; P:nucleosome assembly; IEA.  
 DR InterPro; IPR005819; Histone\_H5.  
 DR PRINTS; PRO0624; HISTONEH5.  
 KW Complete proteome.  
 SQ SEQUENCE 372 AA; 4013 MW; 87649785ECC3C0BC CRC64;

	Query March	41.5%;	Score 88.5;	DB 16;	Length 372;	
	Best Local Similarity	39.7%;	Pred. No. 0.33			
	Matches	27;	Conservative	8;	Mismatches	10; Indels 23; Gaps 1
Cy	1 AKKVKAKKAEEAAEKKKYKAAAEKKAKAYEKA-----AERKA	37				
	:	:	:	:	:	:
Dd	121 AEDMAAGTAEEAAQAAAEKKGALEKKGDDEAKGALIERQQADIAKKRAEDEAKKAAEEEEKKA	180				
Oy	38 AKEAATZA 45					
	:	:	:	:	:	:
Dd	181 AEEAKKKA 188					

Search completed: April 20, 2004, 23:17:44  
Job time : 23.039 secs







Query Match	63.1%;	Score 134.5;	DB 3;	Length 56;
Best Local Similarity	68.4%;	Pred. No. 1.1e-07;		
Matches 39;	Conservative	0;	Mismatches 5;	Indels 13;
				Gaps 4

RESULT 4  
AAy82575  
ID AAy82575 standard; peptide; 77 AA

DT	28-JUL-2000	(first entry)	.
XX			
DE	Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:5		
VV			

KW Copolymer/molecular weight marker; TV-marker/immune disease;  
 KW glutarimide acetate; auto/immune disease; anti/arthritis; neuroprotective;  
 KW osteopathic; immunosuppressive; antithyroid; anti/inflammatory;  
 KW antidiabetic; thyromimetic; haemostatic; antispasmodic; dermatological;  
 KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
 KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
 KW Cohn's disease; chronic immune thrombocytopenia purpura; colitis;  
 KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis  
 KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
 KW pemphigus vulgaris; systemic lupus erythematosus.

OS	Unidentified.
XX	
PN	WO200018794-A1

06-APR-2000.

PF 24-SEP-1999; 99WO-US022402.

PR 25-SEP-1998; 98US-0101693P.

PA (YEDA ) YEDA RES & DEV CO LTD.

1000

DR WPI; 2000-317499/27.

XX Copolymer 1 related polypeptides used as molecular weight markers for

PT glatiramer acetate and for treatment and prevention of immune diseases.

XX

XX

PS Claim 10; Page 14; 72pp; English.

XX

CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular

CC weight TV-marker polypeptides from the present invention. The present

CC invention describes polypeptides (I) for determining the molecular weight

CC of a copolymer (CP), which has an identified molecular weight and an

CC amino acid composition corresponding to the copolymer. The polypeptides

CC of the invention are used as molecular weight markers for glatiramer

CC acetate related tetrapolymers. The polypeptides may also be used for

CC treating and preventing immune diseases in a mammal. Autoimmune diseases

CC which may be treated include either cell-mediated or antibody-mediated

CC diseases. Such diseases include arthritic conditions, demyelinating

CC diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid

CC arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune

CC celiacitis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's

CC disease, chronic immune thrombocytopenia purpura, colitis, contact

CC sensitively disease, diabetes mellitus, Graves disease, Guillain-Barre's

CC syndrome, Hashimoto's disease, idiopathic myxoedema, myasthenia gravis, a

CC psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated-

CC mediated diseases which can be treated include host-versus-graft disease,

CC graft-versus-host disease, and delayed-type hypersensitivity. The

CC polypeptides of the invention have defined molecular weights and physical

cc makes them ideal for use as molecular weight markers  
xx  
sq Sequence 77 AA;

Query Match	61.5%;	Score 131;	DB 3;	Length 77;
Best Local Similarity	50.6%;	Pred. No. 3.6e-07;		
Matches 39;	Conservative	0;	Mismatches 6;	Indels 32;
				Gaps 3

1 AKRYAKKEKAYAKKAETAKKAETAKAYYAKAEAKKKAETAKKAYAKAKAAEKKKYYAAAEAK 6

Db 61 YKABAKAAKEAAYEA 77

## RESULT 5

ID	AA	Y82576	standard; peptide; 86	AA.
1	AA	Y82576	standard; peptide; 86	AA.

AC AAY82576;

DT 28-JUL-2000 (first entry)

DE	Copolymer molecular weight	TV-marker amino acid sequence	SEQ ID NO:6.
DE			

KW Copolymer; molecular weight marker; TV-marker; immune disease;

osteopathic; immunosuppressive; antithyroid; antiinflammatory; KM

antianaemic; immunosuppressive; demyelinating disease; osteoarthritis; KW

KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;

Hashimoto's disease; idiopathic myxoedema; myasthenia gravis; KM

perempitibus vultibus! byscentibus tibus eycuennallobus.  
XX  
XX  
XX

[illegible]

W0200016/94-AL.  
FIN  
XX

06-APR-2000.  
FD  
XX

PP 24-SEP-1999; 99WO-05022402.



PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (EIT- ) EITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACP31694.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 XX Claim 25; SEQ ID NO 55748; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation; (7) identifying a gene in an operon required for  
 CC proliferation; (8) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (9)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pcr\_sequences  
 XX  
 SQ Sequence 428 AA;  
 Query Match 50.9%; Score 108.5; DB 6; Length 428;  
 Best Local Similarity 60.4%; Pred. No. 0.00058;  
 Matches 32; Conservative 2; Mismatches 8; Indels 11; Gaps 2;

QY 1 AKKTA---KKAKEKAKKAYKAAE-----AKKAATKEKAAEKAAEAA 42  
 DB 199 AKKAADADOKKAEAKKAAOEAKKAAEAAMAKKAAEAAMAKKAAEAAMAKKAAEA 251  
 ID AAY82571 standard; peptide; 35 AA.  
 AC AAY82571;  
 XX

DT 28-JUL-2000 (first entry)  
 XX  
 DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO.1.  
 XX  
 XX Copolymer; molecular weight marker; TV-marker; immune diseases;  
 KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
 KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
 KW antidiabetic; thymostatic; haemostatic; antiporiatic; dermatological;  
 KW immunosuppressive; demyelinating disease; osteoarthritis;  
 KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
 KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
 KW diabetes mellitus; Guillain-Barre's syndrome; psoriasis;  
 KW Hashimoto's disease; idiopathic myxedema; myasthenia gravis;  
 KW pemphigus vulgaris; systemic lupus erythematosus.  
 XX  
 OS Unidentified.  
 XX  
 XX WO2000018794-A1.  
 XX  
 PN 06-APR-2000.  
 PD  
 PD 24-SEP-1999; 99WO-US022402.  
 XX  
 PF 25-SEP-1999; 98US-0101693P.  
 PR  
 PR (YEDA ) YEDA RES & DEV CO LTD.  
 PA (TEVA-) TEVA PHARM USA INC.  
 XX  
 PI Gad A, Lis D;  
 PI WPI; 2000-317499/27.  
 DR  
 XX Copolymer 1 related polypeptides used as molecular weight markers for  
 PT glatiramer acetate and for treatment and prevention of immune diseases.  
 PT  
 XX Claim 10; Page 14; 72pp; English.  
 XX  
 CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
 CC weight TV-marker polypeptides from the present invention. The present  
 CC invention describes polypeptides (I) for determining the molecular weight  
 CC of a copolymer (CP) which has an identified molecular weight and an  
 CC amino acid composition corresponding to the copolymer. The polypeptides  
 CC of the invention are used as molecular weight markers for glatiramer  
 CC acetate related tetrapolymers. The polypeptides may also be used for  
 CC treating and preventing immune diseases in a mammal. Autoimmune diseases  
 CC which may be treated include either cell-mediated or antibody-mediated  
 CC diseases. Such diseases include arthritic conditions, demyelinating  
 CC diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid  
 CC arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune  
 CC coprophitis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's  
 CC disease, chronic immune thrombocytopenia purpura, colitis, contact  
 CC sensitivity disease, diabetes mellitus, Graves disease, Guillain-Barre's  
 CC syndrome, Hashimoto's disease, idiopathic myxedema, myasthenia gravis,  
 CC psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated-  
 CC mediated diseases which can be treated include host-versus-graft disease,  
 CC graft-versus-host disease, and delayed-type hypersensitivity. The  
 CC polypeptides of the invention have defined molecular weights and physical  
 CC properties which are analogous to glatiramer acetate molecules, which  
 CC makes them ideal for use as molecular weight markers  
 XX  
 SQ Sequence 35 AA;  
 Query Match 50.7%; Score 108; DB 3; Length 35;  
 Best Local Similarity 64.4%; Pred. No. 5e-05;  
 Matches 29; Conservative 1; Mismatches 5; Indels 10; Gaps 2;

QY 1 AKKTA---KKAKEKAKKAYKAAE-----AKKAATKEKAAEKAAEAA 45  
 DB 1 AKKTAKEKKA--AKKAYK-----KKAKEKAAEAAMAKKAAEA 35  
 ID AAR06446  
 AC AAR06446  
 XX

ID	AA06446	standard; protein; 106 AA.
XX		
AC	AA06446;	
XX		
DT	25-MAR-2003	(revised)
DT	03-JAN-1991	(first entry)
XX		
DE	Recombinant copolymer 1-19, myelin basic protein analogue.	
XX		
KW	Recombinant copolymer 1, COP-1-19; myelin basic protein; MBP;	
KM	immunological activity; autoimmune encephalomyelitis; multiple sclerosis.	
XX		
OS	Synthetic.	
PN	EP83620-A.	
PD	22-AUG-1990.	
XX		
PF	16-FEB-1990; 90BP-00301700.	
PR	17-FEB-1989; 89US-00312541.	
PR	07-FEB-1990; 90US-00473845.	
XX		
PA	(REPK ) REPLIGEN CORP.	
PI	Cook KS;	
DR	WP1; 1990-255848/34.	
DR	N-PDB; AAO06446.	
PT	Producing genes encoding random polymers of aminoacid(s) - for producing recombinant polypeptide(s) with biological and/or immunological activity.	
PS	Disclosure; Fig 12; 25pp; English.	
XX		
CC	To improve the expression of rCOP-1 polypeptides in E. coli, genes coding for rCOP-1-19 were subcloned from pREV 2.1 to pBG3-2deltaN (deposit: 20-Nov-1984 US4691009, NRRL B-15910), a plasmid used to express Protein A. The resulting plasmids encode fusion proteins consisting of beta-glycuronidase, Protein A, and rCOP-1 sequences. A methionine residue occurs between the Protein A and rCOP-1 sequences, originating from the 5' linker sequence, in order that the COP-1 polypeptide may be cleaved from the fusion protein. rCOP-1-19 contains oligonucleotide duplexes encoding the following segments: YKK, AAE, KAK, EKA, KXA, YEA, AYA KEA, CC and AAA. The N-terminal alanine residue is left behind following CNBr cleavage of the fusion protein. The product prevents or arrests experimental autoimmune encephalomyelitis. They are used to prevent, arrest or control a demyelinating disorder. e.g. multiple sclerosis. They may also be used as additives to hair care products to confer beneficial effects on damaged hair or as supplements for diets deficient in certain amino acids. See also AAO05664. (Updated on 25-MAR-2003 to correct PA field.)	
SO	Sequence 106 AA;	
Db		
Oy	3 KYAKKA-KAEKAKAYKAAEAKKAYKAAEKRA-----AKEA 41         :       :       :       :       8 KAKKAYEAKKAKYEAKEAKKAEKAKAKAAEAKKAEAKKAEKAEKAE 55	
Query Match	46.7%; Score 99.5; DB 2; Length 106;	
Best Local Similarity	58.3%; Pred. No. 0.0013;	
Matches	28; Conservative 4; Mismatches 7; Indels 9; Gaps 2	
ID	AA028871	standard; peptide; 46 AA.
AC	AA028871;	
DT	25-MAR-2003	(revised)
DT	23-MAR-1993	(first entry)
XX		

DE	High affinity macrophage mannose receptor ligand compound #9.	
XX	glycopeptide; mannose; mannosylated; glycosylated; mannose receptor;	
KM	macrophages; monocytcs; dectroy; cytotoxicity; label; image; alter;	
KM	macrophage processing of antigen; MHC restriction; inflammation;	
KM	inflammatory diseases; macrophage secretory products; Crohn's disease;	
KM	lymphoma; disease; mononuclear phagocytes; HIV; AIDS;	
KM	lysosomal storage diseases; Gaucher's disease; asthma;	
KM	alveolar macrophages metastasis; systemic macrophages; deliver;	
KM	antigenic peptides; prevent transplant rejection; organ transplantation	
KM	antitumour agents; cancer; toxins.	
XX	Synthetic.	
OS		
XX	Key	Location/Qualifiers
FT	Modified-site	1 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine. May also have non interfering substituents."
FT	Modified-site	4 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	7 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	10 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	13 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	16 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	19 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	22 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	25 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	28 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	31 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	34 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	37 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	40 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	43 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	46 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine."
FT	Modified-site	49 /note= "opt may have mannose, fucose, glucose or N-Ac-glucosamine. May also have non interfering substituents."
XX	WO9219248-A1.	
PN	12-NOV-1992.	
PD	01-MAY-1992.	92MO-US003609.
PF	03-MAY-1991.	91US-00694983.
FR	(UNITW ) UNIV WASHINGTON.	





XX MO20027183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WC-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362659P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KI, Zyskind JW,  
XX Wall D, Trawick JD, Carr GT, Yamamoto R, Foreyth RA, Xu HH,  
XX WPI; 2003-029926/02.  
XX N-PSDB; AQA36489.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 25; SEQ ID NO 60543; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 613 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
XX CC encoding a polypeptide whose expression is inhibited by the antisense  
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX CC polypeptide or its fragment whose expression is specifically binding  
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX CC proliferation or the activity of a gene in an operon required for  
XX CC proliferation; (7) identifying a compound that influences the activity of  
XX CC the gene product or that has an activity against a biological pathway  
XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
XX CC identifying a gene required for cellular proliferation or the biological  
XX CC pathway in which a proliferation-required gene or its gene product lies  
XX CC or a gene on which the test compound that inhibits proliferation of an  
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX CC compound's activity; (11) a culture comprising strains in which the gene  
XX CC product is overexpressed or underexpressed; (12) determining the extent  
XX CC to which each of the strains is present in a culture or collection of  
XX CC strains; or (13) identifying the target of a compound that inhibits the  
XX CC proliferation of an organism. The antisense nucleic acids are useful for  
XX CC identifying proteins or screening for homologous nucleic acids required  
XX CC for cellular proliferation to isolate candidate molecules for rational  
XX CC drug discovery programs, or for screening homologous nucleic acids  
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX CC the target prokaryotic essential genes. Note: The sequence data for this  
XX CC patent did not form part of the printed specification, but was obtained  
XX CC in electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 239 AA;  
XX Query Match 44.4%; Score 94.5; DB 6; Length 239;  
XX Best Local Similarity 56.2%; Pred. No. 0.011;  
XX Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;  
QY 2 KKYAKKAKAKAKAK--AYAAAKK--AKYKAAKAAKAAKAAAYE 44  
DB 124 KAAAEKAEADKKQEDAVKAAAKKQEAEEKAAADKAAAEKAAAE 171  
RESULT 15  
AAR06445  
ID AAR06445 standard; protein; 154 AA.

XX AAR06445;  
XX 25-MAR-2003 (revised)  
XX 03-JAN-1991 (first entry)  
XX Recombinant copolymer 1-77; myelin basic protein analogue.  
XX Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;  
XX immunological activity; autoimmune encephalomyelitis; multiple sclerosis.  
XX Synthetic.  
XX EP83620-A.  
XX 22-AUG-1990.  
XX 16-FEB-1990; 90EP-00301700.  
XX 17-FEB-1989; 89US-00312541.  
XX 07-FEB-1990; 90US-00473845.  
XX (REPK ) REPLIGEN CORP.  
XX Cook KS;  
XX WPI; 1990-255848/34.  
XX N-PSDB; AAQ05664.  
XX Producing genes encoding random polymers of aminoacid(s) - for producing  
XX PT recombinant polypeptide(s) with biological and/or immunological activity.  
XX Disclosure; Fig 11; 25pp; English.  
XX To improve the expression of rCOP-1 polypeptides in *E. coli*, genes coding  
XX CC for rCOP-1-77 were subcloned from pREV 2.1 to pBG3-2deltaN (deposit: 20-  
XX CC NOV-1984 US4691009, NRL B-15910), a plasmid used to express Protein A.  
XX CC The resulting plasmids encode fusion proteins consisting of beta-  
XX CC glucuronidase, Protein A, and rCOP-1 sequences. A methionine residue  
XX CC occurs between the Protein A and rCOP-1 sequences, originating from the  
XX CC 5' linker sequence, in order that the COP-1 polypeptide may be cleaved  
XX CC from the fusion protein. rCOP-1-77 contains oligonucleotide duplexes  
XX CC encoding the following segments: YKK, EAE, KAK, MAK, and AAA. The N-  
XX CC terminal alanine residue is left behind following CNBr cleavage of the  
XX CC fusion protein. The product prevents or arrests experimental autoimmune  
XX CC encephalomyelitis. They are used to prevent, arrest or control a  
XX CC demyelinating disorder, e.g. multiple sclerosis. They may also be used as  
XX CC additives to hair care products to confer beneficial effects on damaged  
XX CC hair or as supplements for diets deficient in certain amino acids. See  
XX CC also AAQ05665. (Updated on 25-MAR-2003 to correct PA field.)  
XX Sequence 154 AA;  
XX Query Match 44.1%; Score 94; DB 2; Length 154;  
XX Best Local Similarity 60.5%; Pred. No. 0.0077;  
XX Matches 26; Conservative 2; Mismatches 9; Indels 6; Gaps 2;  
QY 2 KKYAKKAKAEKAKAYKAAEKAKAKAKYKAAEKAAKAAKAAAYE 44  
DB 105 KKYKKKAKKAKYKK--KAKEAEKA---KAAAEKAKAKAEAYK 141  
Search completed: April 20, 2004, 23:15:46  
Job time : 31.2208 secs





RESULT 4  
US-09-816-982A-5  
; Sequence 5, Application US/05816982A  
; Patent No. US20020115103A1  
; GENERAL INFORMATION:  
; APPLICANT: Gad, Alexander

Db 61 YKAEAAKKAYKAEAAKAAKEAAYEA 86

## RESULT 6



FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO: 56483  
LENGTH: 421  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-10-282-122A-56483

Query Match  
Best Local Similarity 46.0%; Score 98; DB 12; Length 421;  
Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

QY 2 KKYAKKAAEKA--KK--AYKAAEKKAAKTEKAAEKAAYEA 45  
DB 247 KKAEEKAAAEKAAADKKAEEKAAADKKA--AKAAEEKAAAEKAAABA 294

RESULT 10  
US-10-282-122A-59321  
Sequence 59321, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Foreyt, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO: 60543  
LENGTH: 239

PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO: 59321  
LENGTH: 323  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-10-282-122A-59321

Query Match  
Best Local Similarity 44.6%; Score 95; DB 12; Length 323;  
Matches 27; Conservative 4; Mismatches 14; Indels 6; Gaps 1;

QY 1 AKKYAKKAAEKA--KK--AYKAAEKKAAKTEKAAEKAAYEA 45  
DB 92 AAEQKKAABAAAKKAAQCEAEKKAAQCEAAKQAAAEKAAEKAAYEA 142

RESULT 11  
US-10-282-122A-60543  
Sequence 60543, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Foreyt, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO: 60543  
LENGTH: 239















```

QY 2 KKYAKKA---EKAYAKKAKKAKKAYAKKEA----KAYAAEAKKKAKAKKAYAKKEA 54
DB 100 KKAATKATATPKAAAPKKGAVKTKAPKKGKPKSAKKAEPKKGKKAAPK 159
QY 55 KAKKEAYKAEKAYAKKAA---KAEKKEVAAE---AKKAAPK---AYKAEKAAAKE 104
DB 160 AEKKPKAAKDPAAKKTTKAAAKPKAEKKPKAAAPKAEKKPKAAKPKAEKKAKAAKPAKK 219
QY 105 AA 106
DB 220 SA 221

```

## RESULT 15

```

S19113
cgcr-4 protein - Chlamydomonas reinhardtii (fragment)
C/Species: Chlamydomonas reinhardtii
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C/Accession: S19113; S14466
R/Makarchuk, W.W.; Mueller, F.W.; Beck, C.F.
Plant Mol. Biol. 18, 143-146, 1992
A/Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements
A/Reference number: S19113; MUID:92119224; PMID:1731966
A/Accession: S19113
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-265 <NAK>
A/Cross-references: EMBL:X17208; NID:G18136; PIDN:CAA35080.1; PID:G18137
C/Genetic8:
A/Gene: cgcr-4

```

```

Query Match 31.6%; Score 164; DB 2; Length 265;
Best Local Similarity 39.4%; Pred. No. 0.00045;
Matches 50; Conservative 21; Mismatches 38; Indels 18; Gaps 2;

```

```

QY 1 AKKYAKKAEKA-----YAKKAKAKKAKKAYAKKAKKAYAKKAAEAKKAKAA 46
DB 79 AEAKYKAEADAPRAAEAPRAVAEAWMAAEAEAPRAEAWMAAEAEWMAAEAEVAPRAAAQA 138
QY 47 KKYAKKAKKAKKAEKAYKAEKAYAKKAAK---AEKKEVAAEAKKAEKAKKAYKAEKAAKAA 102
DB 139 RVAAEAPRAAAAEAEAPRAAAAEAEAPRAAAAEAEAPRAAAAEAEAPRAAAAEAEAPRAAA 198
QY 103 KEAAVEA 109
DB 199 EAKKAEK 205

```

Search completed: April 20, 2004, 23:18:23  
 Job time : 19.1104 secs





RX MEDLINE=97113525; PubMed=8955385;  
 RA Dennis J.J., Lafontaine E.R., Sokol P.A.;  
 RT "Identification and characterization of the tolQRA genes of  
 RT Pseudomonas aeruginosa.";  
 RL J. Bacteriol. 178:7059-7068 (1996).  
 RN [2]  
 RP REVISIONS TO N-TERMINUS.  
 RA Duan K., Sokol P.A.;  
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAOI;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Steyer C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -1- FUNCTION: Involved in the tonB-independent uptake of proteins (by  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane  
 CC (potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U39558; AAC44660.2; -;  
 DR EMBL; AE004530; AAG04360.1; -;  
 DR PIR; E83525; E83525.  
 DR InterPro; IPR006260; TonB\_C.  
 DR TIGRPFAMS; TIGR01352; tonB\_C term; 1.  
 KW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;  
 KW Complete proteome.  
 FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).  
 FT TRANSSEM 17 37 POTENTIAL.  
 FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).  
 FT DOMAIN 209 216 POLY-ALA.  
 SQ SEQUENCE 347 AA; 37935 MW; EEDDA804A095945 CRC64;  
 Query Match 34.5%; Score 179; DB 1; Length 347;  
 Best Local Similarity 49.1%; Pred. No. 1.9e-05;  
 Matches 54; Conservative 17; Mismatches 29; Indels 10; Gaps 4;  
 QY 6 KKAEEAYAKKAAKKEK-----KAYAKKEAKYAKKAAKKAAYAKKAAKKEAKK 59  
 DB 99 OKLQOQVAAKAAAEQKADARAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKK 158  
 QY 60 AYKAAKAYAAKAEKKEVYAAAEAKKAEAKKAYKAEAKKAAEAYEA 109  
 DB 159 A-EDEAKK--KAAEDAKK--AAEDAKKAAAEAKKAAAEAKKAAVEA 204  
 RESULT 4  
 MS11\_DROHY STANDARD; PRT; 344 AA.  
 AC Q08695;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Axoneme-associated protein ms1101(1).  
 GN MS1101(1).  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.  
 OK NCBI\_Taxid=7224;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=Testis;  
 RX MEDLINE=94200512; PubMed=8150205;  
 RA Neesen J., Buemann H., Heinlein U.A.O.;  
 RT "The Drosophila hydei gene Dms101(1) encodes a testis-specific,  
 RT repetitive, axoneme-associated protein with differential abundance in  
 RT X chromosomal deletion mutant flies.";  
 RL Dev. Biol. 162:414-425 (1994).  
 CC -1- FUNCTION: Possible structural role in the sperm tail. It is  
 CC associated with axonemal structures.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).  
 CC -1- TISSUE SPECIFICITY: Testis. Located in spermatocytes and  
 CC spermatid bundles.  
 CC -1- DOMAIN: The predominant structure is alpha-helical.  
 CC -1- POLYMORPHISM: The number of repeats varies between strains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X73480; CAA51875.1; -;  
 DR PIR; S34153; S34153.  
 DR FlyBase; FBgn0011816; Dhyd/ms101(1).  
 DR GO; GO:0005737; Cytoplasm; IDA.  
 DR GO; GO:0005198; Filamentous molecular activity; IEP.  
 DR GO; GO:0007288; Paxoneme assembly; IEP.  
 KW Sperm; Repeat; Multigene family.  
 FT DOMAIN 58 337  
 FT SEQUENCE 344 AA; 37793 MW; 24C65D2510387E2A CRC64;  
 Query Match 33.8%; Score 175.5; DB 1; Length 344;  
 Best Local Similarity 49.2%; Pred. No. 3.1e-05;  
 Matches 58; Conservative 8; Mismatches 35; Indels 17; Gaps 5;  
 QY 1 AKKAAKKEK---AYAKKAAKKEKAYAKKAAKAYKAAE-----AKKAAKAAK 49  
 DB 65 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 124  
 QY 50 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 106  
 DB 125 CAKAAKKEKA--AKKKCAEAKKKE--AAKKKCAEAKKKEKAAKAAKCAEA 177  
 RESULT 5  
 MS12\_DROHY STANDARD; PRT; 1391 AA.  
 AC Q08656;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Axoneme-associated protein ms1101(2).  
 GN MS1101(2).  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OK NCBI\_Taxid=7224;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=95045538; PubMed=7957199;  
 RA Neesen J., Padmanabhan S., Buemann H.;  
 RT "Randomly arranged repeats of a novel highly charged 16-amino-acid  
 RT motif representing the major component of the sperm-tail-specific  
 RT axoneme-associated protein family Dms101 form extended  
 RT alpha-helical rods within the extremely elongated spermatzoa of



DR InterPro: IPR003216; Linkerhist\_N.  
 DR Pfam: PF00538; linker\_histone; I.  
 DR PRINTS: PR00624; HISTONE5.  
 DR ProDom: PD005373; Linkerhist\_N; 1.  
 DR SMART: SM00526; H15; 1.  
 DR KX Chromosomal protein; Nuclear protein; Multi-gene family.  
 SO SEQUENCE 211 AA; 22169 MW; 9F214581334BBE7A CRC64;

Query Match 32.7%; Score 169.5; DB 1; Length 211;  
 Best Local Similarity 48.7%; Pred. No. 5.1e-05;  
 Matches 55; Conservative 8; Mismatches 41; Indels 9; Gaps 4;

QY 3 KYAKKAE-KAYAKKA-KAAKEKAYAKKAAKAAK-KKAAKEKAYAKKAAK--55  
 DB 68 KLGKKEKSDQKAPDAKAKAKLAKKEKKEKKAASKAKKAKAKSKTTTKV 147  
 QY 56 --AKKAYAKKAYAKKAYAKKAYAKKAAKAAKAAKAAKAAKAAKAAKAA 106  
 DB 148 KPAKAKKAPAKKAPAKKAPAKKAPAKKAPAKKAPAKKAPAKKAPAKKAPAK 200

## RESULT 8

WT3 DROHY STANDARD; PRT; 275 AA.

AC 001395;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Anoxemia-associated protein met101(3).  
 GN MST101(3) OR DHMST101.  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 CC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CC NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

NCBI\_TaxID=7224;

DB 51 AEDVKKCEBAKKAAPAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 106  
 QY 58 KENYAKKAYAKKAYAKKAYAKKAYAKKAAKAAKAAKAAKAAKAAKAAKAA 108  
 DB 107 QEA--AOKKCAELAKKEKKE--AAKCCAAKAAKAAKAAKAAKAAKAAK 153

## RESULT 9

FAU DROME STANDARD; PRT; 668 AA.

AC Q9VGX3; Q9SG18; Q9VGX1; Q9VGX2; Q9Y0F9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Anoxia upregulated protein.  
 GN FAU OR G65544.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CC NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;





156	1-2.
160	1-3.
161	

166	170	1-4.
171	175	1-5.
176	180	1-6.
181	185	1-7.
186	190	1-8.
191	195	1-9.
196	200	1-10.

201	205	1-11.
206	210	1-12.
211	215	1-13.

216	220	1-14.
221	225	1-15.
226	230	1-16.
231	235	1-17.

231	235	1-17.
236	288	6 X 8 AA TANDEM REPEATS OF [EA]-A-A-R-X-A-A-E.
236	243	2-1.

245	2-2
254	2-3
263	2-4
252	
261	
270	

272	279	2-5.
281	288	2-6.
884 AA;	42677 MM;	6E94CBC74294DE8C CRC64;

Parameter	Value
Score	163.5
DB	1
Length	384
Pred. NO.	0.0002
W/matched	40
Tdof	7
Conn	3

YAK-KAEKAYAKKAKAEK----	KAYAKKEAKAYKAEAKKKA,KAERAKKYAKE--AA	54
conservative 12;	misallones 42;	index 1;
		gaps 3;

KEAYKA EAKKYAKA PAKAEKKEVAAA EAKKAEAAKAYKAEAAKAAAEAAAY 107  
AAKEKA AKDPAAKEKA AKDPAAKEKA AKAREMAAKEKA AKDPAAKEEAAR 239

```

:| : | | | | :| :| :| :| :| :| :|
EEAARKAAEEAARKAEEEAARI AEEEARKAAEEEAARKAEEALY 292

```

CELEBRATING 50 YEARS

(Rel. 08, Created)  
(Rel. 08, Last sequence update)  
STANDARD; PRI; 224 AA.

as (Chicken):  
(Rel. 41, Last annotation update)  
11.

Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

DM N° A. 10031;

Robins A.J., Madley L.K., Wells J.R.E.;  
 1 complete set of histone H1 gene complement.  
 PubMed=3597432;

of NMR OF 40-113,  
a complete set of vertebrate protein sequences.  
262: 9656-9663 (1987).

2251; Pubmed=8218199;  
Spens G., Muyldermans S., Segers A., Ramakrishnan V.,  
Hallenka K., Wym L.?

steronuclear two-dimensional NMR studies of the globular protein H1: sequential assignment and secondary structure determination

732:11345-11351(1993).

-----  
 NUCLEAR LOCATION: NUCLEAR.  
 -----  
 BELONGS TO THE HISTONE H1/H5 FAMILY.  
 -----

```

CC -1- FUNCTION: Histones H1 are necessary for the condensation of
CC nucleosome chains into higher order structures.
CC -1- FUNCTION: Oncorhynchon II has antibacterial activity against Gram-
CC positive and Gram-negative bacteria at submicromolar
CC concentrations. Potentially important role in mucosal defense.
CC -1- SUBCELLULAR LOCATION: Nuclear. Oncorhynchon II is secreted.
CC -1- TISSUE SPECIFICITY: Oncorhynchon II is expressed in skin.
CC -1- MASS SPECTROMETRY: MW=7195.3; METHOD=MALDI; RANGE=138-206.
CC -1- MISCELLANEOUS: Is thermostable up to 80 degrees Celsius.
CC -1- SIMILARITY: Belongs to the histone H1/H5 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC used by non-profit institutions as long as its content is in no way com-
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X02624; CAB37646.1; -.
CC DR PIR; A02584; HSTR1R.
CC DR HSSP; P08287; IGHC.
CC DR InterPro; IPR005818; Histone_H1/H5.
CC DR InterPro; IPR005819; Histone_H5.
CC DR Pfam; PF00538; linker_histone; 1.
CC DR PRINTS; PR00624; HISTONEH5.
CC DR SMART; SM00526; H15; 1.
CC KM Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
CC KW Acetylation; Antibiotic.
CC FT INITMET 0
CC FT CHAIN 1 206 HISTONE H1.
CC FT CHAIN 138 206 ONCORHYNCHIN II.
CC FT MOD RES 1 ACETYLATION (BY SIMILARITY).
CC FT DOMAIN 27 100 GLOBULAR.
CC FT SITE 137 138 CLEAVAGE (PROBABLE).
CC FT SITE 137 138
CC SQ SEQUENCE 206 AA; 20672 MW; 72C440798066716C CRC64;
CC -----
Query Match 30.9%; Score 160.5; DB 1; Length 206;
Best Local Similarity 47.7%; Pred. No. 0.00019;
Matches 52; Conservative 9; Mismatches 35; Indels 13; Gaps 4;
Db 105 AKKPKAAAPAAKPKAAKPPAAKPKKVAAKKAAVAAKKSPKAKKPTTPKAAKSPK 164
56 AKKEAYKAEAKKVAKAAKAEKKEYYAAAEAKKAAEAAYVAAEAAKAAE 104
165 VKKPA--AAAK---KAAKSPK--ATKAKPKPAAPKAAKKKAAKPAK 205
RESULT 14
H15_HUMAN STANDARD; PRT; 225 AA.
AC P16401; Q14529;
DT 01-AUG-1990 (Rel. 15, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Histone H1.5 (Histone H1a).
GN HIST1H1B OR H1F5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
RX MEDLINE=90130391; PubMed=2613692;
RA Ohe Y., Hayashi H., Iwai K.;
RT "Human spleen histone H1. Isolation and amino acid sequences of three
RL minor variants, H1a, H1c, and H1d.",
RL J. Biochem. 106:844-857(1989).
RP SEQUENCE FROM N.A.

```





```

RESULT 2
Q8WVK1 PRELIMINARY; PRT; 372 AA.
ID Q8WVK1
AC Q8WVK1
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE TOLA protein.
GN TOLA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX MEDLINE=96198174; PubMed=8626299;
RA Rodriguez-Herva J.J.; Ramos-Gonzalez M.I.; Ramos J.;
RT "The Pseudomonas putida peptidoglycan-associated outer membrane
lipoprotein (PAL) is involved in maintenance of the integrity of the
cell envelope."
RT J. Bacteriol. 178:1699-1706(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX Rodriguez-Herva J.J.;
RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX MEDLINE=96422022; PubMed=8824639;
RA Rodriguez-Herva J.J.; Ramos J.;
RT "Characterization of an OprL null mutant of Pseudomonas putida."
RT J. Bacteriol. 178:5836-5840(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX EMBL; X74218; CAB50780.1;
DR GO; GO:0000786; C:nucleosome; IEA.
DR GO; GO:0005634; C:nucleosome; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006334; F:nucleosome assembly; IEA.
DR InterPro; IPR005819; Histone_H5.
DR PRINTS; PR00624; HISTONEH5.
DR TIGFAMS; TIGR01352; torB Cterm; 1.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;

Query Match 36.7%; Score 190.5; DB 2; Length 372;
Best Local Similarity 50.4%; Pred. No. 9.3e-06;
Matches 59; Conservative 14; Mismatches 31; Indels 13; Gaps 4;

QY 5 AKKAEKAYAKKAKAK-----EKKAYAKKEAKAYKAAE-----AKKAKAEAKKAYKAAE 54
DB 118 AKKAEEDA-AKAEAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAE 176
QY 55 K-AKKEAYKAEKAYKAAEAKKEKAYAAAEKAEAKKAEAKKAEAKKAEAKKAEAKKAE 109
DB 177 KKAABEAKKAAEDAKKKAABEAKKAAEDAKKKAABEAKKAAEDAKKKAABEAKKAAEDA 233

RESULT 3
Q8BN16 PRELIMINARY; PRT; 372 AA.
ID Q8BN16
AC Q8BN16
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Biopolymer transport protein TOLA.
GN TOLA OR ppi221.
OS Pseudomonas putida (strain KT2440).

```

```

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E.; Meinel C.A.; Paulsen I.T.; Dodson R.J.; Hilbert H.;
RA Martins dos Santos V.A.P.; Fouts D.E.; Gill S.R.; Pop M.; Holmes M.;
RA Brinkac L.; Beanan M.; DeBoy R.T.; Daugherty S.; Kolonay J.;
RA Madupu R.; Nelson W.; White O.; Peterson J.; Khouri H.; Hance I.;
RA Chris Lee P.; Holtzapple E.; Scanlan D.; Tran K.; Moazed A.;
RA Utterback T.; Rizzo M.; Lee K.; Kosack D.; Moestl D.; Medler H.;
RA Lauber J.; Stepanovic D.; Hoheisel J.; Stratzel M.; Helm S.;
RA Kiewitz C.; Eisen J.; Timmis K.N.; Duesterhoeft A.; Tsemmler B.;
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."
RT Environ. Microbiol. 4:799-808(2002).
RN [2]
RX EMBL; AB016778; AAN66845.1;
DR TIGR; ppi221;
DR GO; GO:0000786; C:nucleosome; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006334; F:nucleosome assembly; IEA.
DR InterPro; IPR005819; Histone_H5.
DR PRINTS; PR00624; HISTONEH5.
KM Complete proteome.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;

Query Match 36.7%; Score 190.5; DB 16; Length 372;
Best Local Similarity 50.4%; Pred. No. 9.3e-06;
Matches 59; Conservative 14; Mismatches 31; Indels 13; Gaps 4;

QY 5 AKKAEKAYAKKAKAK-----EKKAYAKKEAKAYKAAE-----AKKAKAEAKKAYKAAE 54
DB 118 AKKAEEDA-AKAEAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAE 176
QY 55 K-AKKEAYKAEKAYKAAEAKKEKAYAAAEKAEAKKAEAKKAEAKKAEAKKAEAKKAE 109
DB 177 KKAABEAKKAAEDAKKKAABEAKKAAEDAKKKAABEAKKAAEDAKKKAABEAKKAAEDA 233

RESULT 4
Q8T5C8 PRELIMINARY; PRT; 1866 AA.
ID Q8T5C8
AC Q8T5C8
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE MaebL.
GN MAEBL.
OS Plasmodium vivax.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5855;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Salvador;
RA Michon P.; Stevens J.R.; Kaneko O.; Adams J.H.;
RT "Evolutionary relationships of conserved cytosine-rich motifs in
RT adhesive molecules of malaria parasites."
RT Mol. Biol. Evol. 0:0-0(2002).
RN [2]
RX EMBL; AY042083; AAL10508.1;
DR InterPro; IPR008602; Duffy_binding.
DR Pfam; PF05424; Duffy_binding; 1.
SQ SEQUENCE 1866 AA; 212420 MW; DC692D7CFAE7D93F CRC64;

Query Match 36.7%; Score 190.5; DB 5; Length 1866;
Best Local Similarity 49.6%; Pred. No. 4.1e-05;
Matches 61; Conservative 17; Mismatches 22; Indels 23; Gaps 6;

QY 5 AKKAEKAYAKKAKAK-----EKKAYAKKEAKAYKAAEAKKAEAKKAEAKKAEAKKAE 50
DB 1262 AKKAE--AKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAE 1319

```

AP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21531948; PubMed=11677699;  
RA McClelland M., Sanderson K.B., Spieth J., Clifton S.W., Latreille P.,  
Courtney L., Porwollik S., Alt J., Dante M., Du P. Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvoney E.,  
RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;

```

Qy 1 AKYAKKAE---KAYAKKAKAEKKYAKKEAYYAAADAKKAKAEKCTAK-EPAK 55
    :::::  :  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 120 AEBAAQOAELOKOKOEAAAKAADAKKAEEDDKA--AEBAAKKAAADAKKKAEEPAK 177
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Qy 56 AKKEAY-KAEAKYAKKAKAEKKEEAAAEADK--AEAAKYAEAAKAAAEKAA 106
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

Query Match	Best Local Similarity	Score	DB	Length
Matches 63; Conservative 12; Mismatches 28; Indels 18; Gaps 7;	34.3%;	52.1%;	Pred. No. 6,8e-05;	
2 KKYAKKAEKAYAKKAKAKAKK-----AAKKEA-KAYKA-AEAKKAKAEAKKYAKEA 53	145 KQAAEQCKIAAAVAKAKKEQKQAEETAQAQKAEADIVTAQAEQKAEAEAKKEAAVA 204	54 AKAKGEAYKAEAKKYAKAK-----AEKKEVAAAEAKKEAAK-AVKAEAAKAAAKEA 105	205 AAARKQA-DADAKKAEVAAEKAAADAADAEKKAADAE-KKAAAKKVA/AAAAAEAKKKAALAE 262	















ID ABU27824 standard; protein; 428 AA.  
 AC ABU27824;  
 DT 19-JUN-2003 (first entry)  
 DE Protein encoded by Prokaryotic essential gene #13351.  
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 OS Enterobacter cloacae.  
 PN WO200271183-A2.  
 PD 03-OCT-2002.  
 PF 21-MAR-2002; 2002WO-US009107.  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 PA (ELIT-) ELITRA PHARM INC.  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA31694.  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS Claim 25; SEQ ID NO 55748; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences  
 XX  
 XX Sequence 428 AA;  
 Query Match 36.6%; Score 190; DB 6; Length 428;

Best Local Similarity 50.8%; Pred. No. 3.4e-08;  
 Matches 64; Conservative 8; Mismatches 32; Indels 22; Gaps 5;  
 QY 1 AKKTA-----KAEKAYAKKAAKAEKAYAKKAAKAYAAEAKKKAKAAE-----K 47  
 DB 167 AKKAAADAKKAEAAKAAKAAADAKKAEAAKAAKAA--AAADAKKAEAAKAAKAAQEA 224  
 QY 48 KAKAEAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 100  
 DB 225 KAAAEAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 284  
 QY 101 AAKKAA 106  
 DB 285 AAKKAA 290  
 RESULT 8  
 ID ABU28559  
 AC ABU28559 standard; protein; 421 AA.  
 XX  
 AC ABU28559;  
 DT 19-JUN-2003 (first entry)  
 DE Protein encoded by Prokaryotic essential gene #14086.  
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 OS Escherichia coli.  
 PN WO200271183-A2.  
 PD 03-OCT-2002.  
 PF 21-MAR-2002; 2002WO-US009107.  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 PA (ELIT-) ELITRA PHARM INC.  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA32429.  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS Claim 25; SEQ ID NO 56483; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences  
XX  
SQ Sequence 421 AA:  
  
Query Match 36.5%; Score 185.5; DB 6; Length 421;  
Best Local Similarity 53.0%; Pctd. No. 3.7e-08;  
Matches 61; Conservative 11; Mismatches 32; Indels 11; Gaps 5  
  
Qy 1 AKTKAKKAAE---KAYAKKAKAAKEKKYAKKKAAYKAAEAKKKAKAAKVKAK-EAAK 55  
Db 120 AEEAAKAAEIKKQKQAEBAAKAAADADKAAEDADK--AEBAKKAADADKKKAEBAK 177  
Qy 56 AKKEAY-KAEAKKYAKAAAEKKEVAAAEAKK--AEAAKAYKQEAAYAAAKKAA 106  
Db 178 AAEEAQKAAEAAAALAKKKAAEAAEAAAEAAAEARKKAAEAAEKAAEKAALAEKAA 232  
  
RESULT 9  
AAR06446  
ID AAR06446 standard; protein; 106 AA.  
XX  
AC AAR06446;  
XX  
DT 25-MAR-2003 (revised)  
DT 03-JAN-1991 (first entry)  
XX  
DE Recombinant copolymer 1-19, myelin basic protein analogue.  
XX  
XX Recombinant copolymer 1; COP-1-19; myelin basic protein; MBP;  
XX immunological activity; autoimmune encephalomyelitis; multiple sclerosis.  
XX  
XX Synthetic.  
OS  
XX EP283620-A.  
XX  
XX 22-AUG-1990.  
XX  
XX 16-FEB-1990; 90EP-00301700.  
XX  
XX 17-FEB-1989; 89US-00312541.  
XX  
XX 07-FEB-1990; 90US-00473845.  
XX  
XX (REPK ) REPLIGEN CORP.  
PA  
XX  
XX Cook KS;  
XX  
XX WPI; 1990-255848/34.  
XX  
XX N-PSDB; AAO06446.  
XX  
XX Producing genes encoding random polymers of aminoacid(s) - for producing  
XX recombinant polypeptide(s) with biological and/or immunological activity.  
XX  
XX  
XX Disclosure; Fig 12; 25BP; English.  
XX  
XX To improve the expression of rCOP-1 polypeptides in *E. coli*, genes coding  
XX for rCOP-1-19 were subcloned from pREV 2.1 to pBG3-2delEan (deposit: 20-  
XX NOV-1984 USA691009, NRRL B-15910), a plasmid used to express Protein A.  
XX The resulting plasmids encode fusion proteins consisting of beta-  
XX glucuronidase, Protein A, and rCOP-1 sequences. A methionine residue  
XX occurs between the Protein A and rCOP-1 sequences, originating from the

CC	5' linker sequence, in order that the COP-1 polypeptide may be cleaved
CC	from the fusion protein. rCOP-1-19 contains oligonucleotide duplexes
CC	incoding the following segments: YKR, AAE, KAR, EKA, KXA, YEA, AKK XEA,
CC	and AAA. The N-terminal alanine residue is left behind following CNBr
CC	cleavage of the fusion protein. The product prevents or arrests
CC	experimental autoimmune encephalomyelitis. They are used to prevent,
CC	arrest or control a demyelinating disorder, e.g. multiple sclerosis. They
CC	may also be used as additives to hair care products to confer beneficial
CC	effects on damaged hair or as supplements for diets deficient in certain
CC	amino acids. See also AAQ05664. (Updated on 25-MAR-2003 to correct PA
CC	field.)
CC	
SQ	Sequence 106 AA:
Query Match	36.0%; Score 187; DB 2; Length 106;
Best Local Similarity	55.0%; Pred. No. 1,4e+08;
Matches	60; Conservative 11; Mismatches 20; Indels 18; Gaps 6;
OY	5 AKKAEEKAYAKKAKAEEKKAYAKKAAVYAAEAKKAKAEKKAYAKEAAKAEYAAKE 64
DB	1 AKAEKKAKAK-KAYEEKAKAKAYEAK--KAEKEKAEKAAEKKAEEAKKAEK----- 51
OY	65 AKKTAAKAAEKKEKYAAAEKKKAEEA---KAYAAEAAKAAK-EAAVE 108
DB	52 AKAEKAKAEKAEK---AKAEKAAVAAAEKAEKAEKAEKAEKAAKAAK 96
RESULT 10	
ID.	ABU47123
XX	ABU47123 standard; Protein: 407 AA.
AC	
XX	ABU47123;
XX	
DT	19-JUN-2003 (first entry)
DE	Protein encoded by Prokaryotic essential gene #32650.
XX	
KM	Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX	
OS	Salmomella typhimurium.
PN	
XX	WO200277183-A2.
PD	
XX	03-OCT-2002.
PF	
XX	21-MAR-2002; 2002W0-USO09107.
PR	
XX	21-MAR-2001; 2001US-00815242.
PR	06-SEP-2001; 2001US-00948993.
PR	25-OCT-2001; 2001US-0342933P.
FR	08-FEB-2002; 2002US-0072851.
XX	06-MAR-2002; 2002US-0326599P.
PA	(ELIT-) ELITRA PHARM INC.
PI	
XX	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI	Wall D, Treweek JD, Carr GT, Yamamoto R, Foreyth RA, Xu HH;
DR	
XX	WP1; 2003-028926/02.
DR	N-PADB; ACA50993.
PT	
XX	New antisense nucleic acids, useful for identifying proteins or screening
PT	for homologous nucleic acids required for cellular proliferation to
PT	isolate candidate molecules for rational drug discovery programs.
XX	
PS	Claim 25; SEQ ID NO 75047; 1766bp; English.
XX	
CC	The invention relates to an isolated nucleic acid comprising any one of
CC	the 623 antisense sequences given in the specification where expression
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:
CC	(1) a vector comprising a promoter operably linked to the nucleic acid
CC	encoding a polypeptide whose expression is inhibited by the antisense
CC	nucleic acid;(2) a host cell containing the vector; (3) an isolated



[illegible]

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification, where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from Wipo at [http://wipo.int/publ/published\\_pct\\_sequences](http://wipo.int/publ/published_pct_sequences)

QY 1 AKKAYKKEKAYAKKAAKKEKAYAKKEAAYAAAKKKAKKEAKKYKAEAK-----55  
 Db 47 ADAGAAAEQAARAAADAK-KQAE---AAKAAAEKQAEKAAKAAAEQKAAEA 102  
 QY 56 -AKKEAYKAEKAYAKAAKAEKAYAAAEKAAKAAAYAYAAAKAAKAAKAA 106  
 103 AAKKQOEKKAQOEAAKQAAAEAAAE-QAAKAAAYQAARAAKAAAEKAA 153

KW diabetes mellitus: Graves disease: Guillain-Barre's syndrome  
 KW chronic disease: chronic immune thrombocytopenia purpura; colitis;  
 KW diabetes mellitus: Graves disease: Guillain-Barre's syndrome





[illegible]



```

; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68109
;
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Pseudomonas putida
;
; US-10-282-122A-68109
;
Query Match          36.7%; Score 190.5; DB 12; Length 372;
Best Local Similarity 50.4%; Pred. No. 9.1e-08;
Matches 59; Conservative 14; Mismatches 31; Indels 13; Gaps 4;

QY 5 AKKAAYAKKAKKAAK-----EKKAYAKKAKAYKAAE-----AKKKAKAKKAYAKAA 54
DB 118 AKKAEDA-AKAAEAABAKAAAKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAK 176
QY 55 K-AKKAAYAKKAKKAKKAAKAAEKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAK 109
DB 177 KKAABEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAK 233

RESULT 6
US-10-282-122A-55748
; Sequence 55748, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
```

```

; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55748
;
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
;
; US-10-282-122A-55748
;
Query Match          36.6%; Score 190; DB 12; Length 428;
Best Local Similarity 50.8%; Pred. No. 1.2e-07;
Matches 64; Conservative 8; Mismatches 32; Indels 22; Gaps 5;

QY 1 AKKA-----KKAAYAKKAKKAKKAYAKKAAKAYKAAEAKKAAE-----K 47
DB 167 AKKAADAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAK 224
QY 48 KYAKEAAKAKKAYKAAEAKKAYKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAK 100
DB 225 KAAAEAAKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAA 284
QY 101 AKKAA 106
DB 285 AAKKAA 290

RESULT 7
US-10-282-122A-56483
; Sequence 56483, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
```

```

QY      1 AAKYAKKAEEAVKKKKAAKKKKAAKKKAAKAAKVK--AAEKKKAAKKAEEAKYAAEA--AKKK 58
Dd      138 AKAAADAKKAAEEAAKAAADAK--KKAEEAVKAAADAKKAAEEAAKAAADAKKKEA 19
QY      59 EAVK--AAEAKYAKAAKAAKEKAYAAAEAK--AAEAKAYKAAK--AAAKEAYEA 109
Dd      196 EAAKAAAEAAKKAEEAAK---AAEEAAKKKDAEAAK--AAEAKKADAAAKAAEA 249

RESULT 9
US-10-282-122A-59321
; Sequence 59321, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl

```

```

APPLICANT : Zyskind, Judith
APPLICANT : Wall, Daniel
APPLICANT : Trawick, John
APPLICANT : Carr, Grant
APPLICANT : Yamamoto, Robert
APPLICANT : Forsyth, R.
APPLICANT : Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 59321
LENGTH: 323
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-10-282-122A-59321

Query Match      35.0%; Score 181.5; DB 12; Length 323;
Best Local Similarity 52.7%; Pred. No. 4,1e-07;
Matches 59; Conservative 8; Mismatches 34; Indels 11; Gaps

1 AKYKAAKAEKAYKAKKAKKEKKYAKKAEKAYKAAAEKKKAAAKAKKYAKAAK----- 55
Db 47 ADQAQAAQAAQAAKAAKAAKAAK--KQAEK---AAAKAAAEAKQKQAEAPAAKAAEAQKKAERA 106
56 -AKKEAYKAEAKYKAYKAAAEKKEEYAAAEEAKKAEBAKAYKKEAAKAAKAAKAA 106

```

D6 103 AAKKAGQGEAKKAGQEAAGAAEKAAE-KAAEKAAQKAAAEKAAEKAA 153

[illegible]

```

PRIORITY APPLICATION NUMBER: 60/257,991
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 66237
LENGTH: 347
TYPE: PRN
ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66237
```

	Query Match	34.5%	Score 179; DB 12; Length 347;
	Best Local Similarity	49.1%;	Pred. No. 6.9e-07;
Matches	54; Conservative	17; Mismatches	29; Indels 10; Gaps 4
Oy	6 KKAERKAAKAAAEK-----KAYAKKEKKYKAAEAKKKAAEKKYAAEAAYAKE	59	
Db	99 QRLGGQGVAAAAGAQKADARFAAENQAABAAKKNDDEAKCAAFAPAAAEQGKQADAKRR	158	
Oy	60 AYKAEKKYAKAAAKAEKYEVAALAAAKQAEEPAKAYKAAALAAKAEAYEA	109	
Db	159 A-EDEAKRKK--RAEDARKRK-AAEDPARKKAAEAAEQGAAMEAPAKKAAVVEA	204	

```

RESULT 12
US-10-127-032-120
Sequence 120, Application US/10127032
Publication No. US20030113742A1
GENERAL INFORMATION:
APPLICANT: Whiteley, Marvin
APPLICANT: Bangera, M. Gita
APPLICANT: Lory, Stephen
APPLICANT: Greeberg, Everett Peter
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
FILE REFERENCE: UIZ-0700P
CURRENT APPLICATION NUMBER: US/10/127,032
CURRENT FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: US 60/285,190
PRIOR FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: US 60/344,142
PRIOR FILING DATE: 2001-10-24
NUMBER OF SEQ ID NOS: 170
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 120
LENGTH: 347
TYPE: RPT
ORGANISM: Pseudomonas aeruginosa
US-10-127-032-120

```

Query Match	34.5%;	Score 179;	DB 14;	Length 347;
Best Local Similarity	49.1%;	Pred. No. 6,9e/07;		
Matches	54;	Conservative 17;	Mismatches 29;	Indels 10;
			Gaps	4;
QY	6	KKAEEAYKKKKKKKKKK-----KAYKKKKKKKKYKAAEKKKKAAKKYAAEAAKKKE	59	
DB	99	QTELEQQYAAAKAAEKKKKADDEKARAKAEKKAAEKKKKDDEAKKAAEAAKAAEKKKKADITAKK	158	
QY	60	ATKAEAKTYAAQAATKAEKKKYAAAEAKKAAEAAKAYKAAKAAKAAEAYAA	109	
DB	159	A-EDAAKK--KAAEDAKKK--AAEDAKKKAAEKKKKAAAEAAKCKAAAYAA	204	

RESULT 13  
US-10-051-643-201  
; Sequence 201, Application US/10051643  
; Publication No. US20020197265A1  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.



```

1  APPLICANT: Tan, Paul L. J.
2  TITLE OF INVENTION: Methods and Compounds for the Treatment
3  TITLE OF INVENTION: of Immunologically-Mediated Diseases of the Respiratory
4  TITLE OF INVENTION: System using Mycobacterium Vaccae
5  FILE REFERENCE: 11000.1008C2
6  CURRENT APPLICATION NUMBER: US/10/051,643
7  CURRENT FILING DATE: 2002-01-18
8  PRIOR APPLICATION NUMBER: US09/156,181
9  PRIOR FILING DATE: 1998-09-17
10 PRIOR APPLICATION NUMBER: US 08/996,624
11 PRIOR FILING DATE: 1997-12-23
12 NUMBER OF SEQ ID NOS: 208
13 SOFTWARE: FastSeq for Windows Version 3.0
14 SEQ ID NO: 201
15 LENGTH: 223
16 TYPE: prt
17 ORGANISM: Mycobacterium vaccae
18 US-10-051-643-201

```

Query Match	34.3%;	Score 178;	DB 13;	Length 223;
Best Local Similarity	54.5%;	Pred. No. 5.2e-07;		
Matches	60;	Mismatches	34;	Indels 10; Gaps 5

```

QY      1 AKTKAKK--AEKAYAKKAAKE---KKAYAKKEAKAYKAAEAKKKAKKAYAKKEAAK 55
      ||| : ||| ||| : ||| ||| ||| : ||| ||| ||| : |||
DB     112 ARKAAKKAPAKKAAAKKAPAKKAPAKKA-ATKAPAKKATAAKGAAPAKKATAAKKAAP 170

```

QY 56 AKKEAYKAEAKKYAKAAKAKEKEYA--AAEAKKAEEAKAYKKYEAAKAAAK 103  
| | | | | : | | | | |  
Db 171 AKKAPAKKAATTAAPAKAPAKKAATTAAPAKKAPAK--KAPAKKAPAK 218

```

RESULT 14
US-10-205-979-52
Sequence 52, Application US/10205979
Publication No. US20030147861A1
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Tan, Paul L. J.
TITLE OF INVENTION: Compounds and Methods for the Modulation of Immune Responses
FILE REFERENCE: 11000.1063U
CURRENT APPLICATION NUMBER: US/10/205,979
CURRENT FILING DATE: 2002-07-25
PRIOR APPLICATION NUMBER: 60/308,446
PRIOR FILING DATE: 2001-07-26
NUMBER OF SEQ. ID NOS: 52
SOFTWARE: FastSeq for Windows Version 4.0
SEQ. ID NO: 52
LENGTH: 223
TYPE: PRT
ORGANISM: Mycobacterium vaccae
US-10-205-979-52

```

[illegible]

```

RESULT15
US-10-282-122A-75772
; Sequence 75772, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:

```

```

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EPITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/151,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 75772
LENGTH: 376
TYPE: PRT
ORGANISM: Salmonella typhi
US-10-282-122A-75772

```

Query Match	34.3%	Score 178;	DB 12;	Length 376;
Best Local Similarity	52.3%	Pred. No. 9e-07;		
Matches	58;	Conservative	10;	Mismatches 33; Indels 10; Gaps 4

  

Cy	1	AKKVAKKKEKVAKKAKKAKKVAKKKKKAKKAAKKA	60
Db	137	AKAAADAKKKAAEAARAAADAKKKAEBA-AKAAADAKKKAAEA--AKAAAEKK--	190
Cy	61	YKAQKCAKAAAKKEKYYAAAEKKADAAKAYAEAAKAA--KEAYE	108
Db	191	-KAEAAAGAAADAKKKDAEAATAAEAKKKDAAAAAKAAADAKKKAEE	240

Search completed: April 20, 2004, 23:27:18  
Job time : 54.7922 secs

RESULT 15

```

; Sequence 75772, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:

```

**ପାଞ୍ଚ ୬**

```

1 FILE REFERENCE: 2709.1002-001
2 CURRENT APPLICATION NUMBER: US/09/543,681A
3 CURRENT FILING DATE: 2000-04-05
4 PRIOR APPLICATION NUMBER: US 60/128,706
5 PRIOR FILING DATE: 1999-04-09
6 NUMBER OF SEQ ID NOS: 8344
7 SEQ ID NO 5390
8
9 LENGTH: 361
10 TYPE: PRT
11 ORGANISM: Proteus mirabilis
12
13 US-09-543-681A-5350

```

Query Match	32.6%;	Score 169;	DB 4;	Length 361;
Best Local Similarity	44.0%;	Pred. No. 3,2e-07;		
Matches 51;	Conservative 16;	Mismatches 37;	Indels 12;	Gaps 2

QY 1 AKTKAKAAEAYVKKAAALAEKKYAVKAEAKAYKA-----AEAKKKKAAEKYAK 51  
Db 123 AKEMAEQREAEATAAAKAKEEQKQAEEMAAQAKMERDRITLKEQDAIVAKQAEAKKQAE 189  
QY 52 EAAYAKAEAYKAEKKYAKA---AKKEKEVYAAAEKKYAAKAYKAEAAATAAAE 104  
Db 163 LAAKQKQAEAKQAEEAEAKQAEADAKAEADAKAEADAKAAAEAKQAEEAKQAEEAKQAEE 238

RESULT 15  
US-08-216

Sequence 8, Application US/08216894  
Patent No. 5876734  
GENERAL INFORMATION:  
APPLICANT: Kirchhoff, Louis V.  
APPLICANT: Otsu, Keiko  
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
TITLE OF INVENTION: WITH TRYPANOSOMA CRUZI  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Releasee #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/216,894  
FILING DATE: 24-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5395  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 643 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-216-894-8

Query Match	31.3%;	Score 162.5;	DB 2;	Length 643;
Best Local Similarity	44.4%;	Pred. No. 2e-06;		
Matches 52;	Conservative 18;	Mismatches 36;	Indels 11;	Gaps 4

QY I AKTKAKKAEEAYAKUKAKAKEKAYAKKEAKYKAAEEKKAKAAEAKKYA----KEAAKA 56  
 D5 445 AAEEATKVAEEAEKQKAAEATKVAEEAEKQKAAEATKVAEEAEKQKAAEATKVAEEAEKQKAAEA 504

```

Oy      57 KK--EAYKBAEKVYKAAKAEKEEYA---AAEAKKBAEAKVYK-AEAAKAAAEAA 106
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      505 TKVAEAEKQKAAEATVVAEAEKQKAAEATKVAAEAEKQKAAEATVVAEAEKQKAAEAA 561

```

Search completed: April 20, 2004, 23:19:08  
Job time : 23.6494 secs

Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc C. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 29581  
LENGTH: 407  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29581

Query Match 34.5%; Score 179; DB 4; Length 407;  
Best Local Similarity 49.1%; Pred. No. 5.2e-08;  
Matches 54; Conservative 17; Mismatches 29; Indels 10; Gaps 4;

QY 6 KKAERKAYAKKAKAKK-----KAYAKKAYKAAKAKKAYKAAKAKKAYKAAKAKK 59  
DB 159 QKLEOQOQVAAAKAAEQKDEARKAAEQKAAKAAKAAKAAKAAKAAKAAKAAKAAK 218  
QY 60 AYAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 109  
DB 219 A-EDEAKK-KAAEDAKK-KAEDAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 264

## RESULT 12

US-09-095-855-201  
Sequence 201, Application US/09095855

Patent No. 6160093  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: Compounds and Methods for  
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
NUMBER OF SEQUENCES: 208  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,855  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/705,347  
FILING DATE: 29-AUG-1996  
APPLICATION NUMBER: 08/873,970  
FILING DATE: 12-JUN-1997  
APPLICATION NUMBER: 08/997,362  
FILING DATE: 23-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002c3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565

TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 201:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-095-855-201

Query Match 34.3%; Score 178; DB 3; Length 223;  
Best Local Similarity 54.5%; Pred. No. 3.5e-08;  
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;

QY 1 AKKYAKK--AEKAYAKKAKAKE--KAYAKKAYKAAKAAKAAKAAKAAKAAKAAKAAK 55  
DB 112 ARKAAKAPAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 170  
QY 56 AKKAYKAAKAYKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 103  
DB 171 AKKAPAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 218

## RESULT 13

US-09-205-426-201  
Sequence 201, Application US/09205426

Patent No. 6406704  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
APPLICANT: Tan, Paul L. J.  
TITLE OF INVENTION: Compounds and Methods for Treatment and  
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
FILE REFERENCE: 11000.1002c4  
CURRENT APPLICATION NUMBER: US/09/205,426  
CURRENT FILING DATE: 1998-12-04  
EARLIER APPLICATION NUMBER: 09/095,855  
EARLIER FILING DATE: 1998-06-11  
EARLIER APPLICATION NUMBER: 08/997,362  
EARLIER FILING DATE: 1997-12-23  
EARLIER APPLICATION NUMBER: 08/873,970  
EARLIER FILING DATE: 1997-06-12  
EARLIER APPLICATION NUMBER: 08/705,347  
EARLIER FILING DATE: 1996-08-29  
NUMBER OF SEQ ID NOS: 208  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 201  
LENGTH: 223  
TYPE: PRT  
ORGANISM: Mycobacterium vaccae  
US-09-205-426-201

Query Match 34.3%; Score 178; DB 4; Length 223;  
Best Local Similarity 54.5%; Pred. No. 3.5e-08;  
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;

QY 1 AKKYAKK--AEKAYAKKAKAKE--KAYAKKAYKAAKAAKAAKAAKAAKAAKAAKAAK 55  
DB 112 ARKAAKAPAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 170  
QY 56 AKKAYKAAKAYKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 103  
DB 171 AKKAPAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 218

## RESULT 14

US-09-543-681A-5390  
Sequence 5390, Application US/09543681A

Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABI  
DIAGNOSTICS AND THERAPEUTICS







